

SEARCH REQUEST FORM

6-963

Examiner # (Mandatory): _____ Requester's Full Name: Li LeeArt Unit 1645 Location (Bldg/Room#): 8E17 Phone (circle 305 306 308) 8891Serial Number: 09/041,236 Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Keywords (include any known synonyms registry numbers, explanation of initialisms):

Please search

- (1) each residues (in claims 11 and 12) of SEQ ID NO=2
see the attached claims
- (2) 1-190.2 of SEQ ID NO=1 (NA)

Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

Thanks

STAFF USE ONLY

Searcher: D. SchuberSearcher Phone #: 302-4292Searcher Location: CM1 12F18Date Picked Up: 7/8Date Completed: 7/12

Clerical Prep Time: _____

Terminal Time: 4Number of Databases: 7

Type of Search

1 N.A. Sequence4 A.A. Sequence

____ Structure (#)

____ Bibliographic

____ Litigation 1

____ Fulltext

____ Procurement

____ Other

Vendors (include cost where applicable)

____ STN

____ Questel/Orbit

____ Lexis/Nexis

____ WWW/Internet

✓ In-house sequence systems (list) MSU/L
CompuGen

____ Dialog

____ Dr. Link

____ Westlaw

____ Other (specify)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 1999, 18:47:21 : Search time 905.34 seconds
(without alignments)
7515.957 Million cell updates/sec

Title: US-09-041-236-1_copy_1_1902

Perfect score: 1902

Sequence: 1 CTGCTGCTGCTGCTGCTGGC.....CTCTGCTGCTGCTGCTGCTCAC 1902

Scoring table: IDENTITY_NUC

Searched: 808301 seqs, 1788773984 residues

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl1.*
- 9: gb_pl2.*
- 10: gb_pr1.*
- 11: gb_pr2.*
- 12: gb_pr3.*
- 13: gb_ro.*
- 14: gb_st.*
- 15: gb_sy.*
- 16: gb_un.*
- 17: gb_vi.*
- 18: gb_htg.*
- 19: em_ba.*
- 20: em_fun.*
- 21: em_hum1.*
- 22: em_hum2.*
- 23: em_in.*
- 24: em_om.*
- 25: em_or.*
- 26: em_ov.*
- 27: em_pat.*
- 28: em_ph.*
- 29: em_pl.*
- 30: em_ro.*
- 31: em_sy.*
- 32: em_un.*
- 33: em_vi.*
- 34: em_htg.*
- 35: em_sts.*
- 36: gb_ba1.*
- 37: gb_ba2.*
- 38: gb_pl1.*
- 39: gb_pl2.*
- 40: gb_pr1.*
- 41: gb_pr2.*
- 42: gb_pr3.*
- 43: gb_sts.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Query	Description
-----	-----	-----	-----	-----	-----	-----	-----

1	1902	100.0	2661	12	AF069493
2	1902	100.0	2594	12	AF071542
3	1902	100.0	2661	42	AF069493
4	1902	100.0	2594	42	AF071542
5	1897.2	99.7	2636	12	AF030698
6	1897.2	99.7	2636	42	AF030698
7	853.8	44.9	1192	13	AF030699
8	380	20.0	130608	17	AF005370
9	380	20.0	11874	17	AHU18243
10	366	19.2	8861	12	AF030697
11	366	19.2	8861	42	AF030697
12	70.6	3.7	3552	10	HSU38276
13	70.6	3.7	2719	12	HSU33920
14	70.6	3.7	3552	40	HSU38276
15	70.6	3.7	2719	42	HSU33920
16	68	3.6	2383	13	AF080090
17	68	3.6	2265	13	AF080091
18	63.2	3.3	2919	10	HSU28369
19	63.2	3.3	2919	40	HSU28369
20	59	3.1	2770	13	AF030430
21	58.2	3.1	2993	5	AF073289
22	57.8	3.0	2769	13	MM069535
23	57.6	3.0	5177	11	AB000220
24	57.6	3.0	5177	41	AB000220
25	56.4	3.0	2336	5	AF022946
26	55.2	2.9	36552	11	HSU90094
27	55.2	2.9	36534	11	U73167
28	55.2	2.9	36552	41	HSU90094
29	55.2	2.9	36534	41	U73167
30	54.2	2.8	2872	13	MMRNASEMA
31	53.8	2.8	4157	12	HSU60800
32	53.8	2.8	4157	42	HSU60800
33	52.8	2.8	2356	13	MUSC1
34	52.8	2.8	2913	13	MMRNASEMD
35	52.6	2.8	3503	13	S79463
36	51.2	2.7	1998	13	MUSSEMAIII
37	51.2	2.7	2477	13	MMRNASEME
38	50	2.6	2433	13	MMRNASEMC
39	49.4	2.6	8056	11	HSU52840
40	49.4	2.6	8056	41	HSU52840
41	48.4	2.5	3847	13	AB000776
42	47	2.5	2715	5	GGU28240
43	46.8	2.5	6914	17	HSVUSRORFS
44	46.4	2.4	3635	13	AF036585
45	46	2.4	2396	10	HSBCAM

ALIGNMENTS

RESULT 1

LOCUS	AF069493	2661 bp	mRNA	PRI	03-DEC-1998
DEFINITION	Homo sapiens GPI-anchored membrane protein CDw108 precursor, mRNA, complete cds.				
ACCESSION	AF069493				
NID	g3947939				
KEYWORDS	human.				
SOURCE	Homo sapiens				

ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2661)				
AUTHORS	Yamada, A., Kubo, K., Takeshita, T., Harashima, N., Kawano, K., Sagawa, K., Sugamura, K. and Itoh, K.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1998) Immunology, Kurume University School of Medicine, Asahi-machi 67, Kurume, Fukuoka 830, Japan				
FEATURES	Location/Qualifiers				
source	1..2661				
	/organism="Homo sapiens"				
	/db_xref="taxon.9606"				
	/chromosome="15"				

```
/map="between WI-6247 and CHLC.GATA85D02"
/cell_line="HPB-ALL"
/cell_type="T lymphocyte"
18..2018
/notes="putative adhesion molecule; lymphocyte activation
antigen"
/codon_start=1
/product="GPI-anchored membrane protein CDw108 precursor"
/db_xref="PID:g3947940"
/translation="MTPTPPGAAAPSAPRARVPGPPARLGLRLRLRLLLWAAASA
QHLRSGRIFAWKGVHGDVDFQTEPHTVLHEPGSSVWVYGRKXVLFDPPE
GKNSVRYTNGISTKSCLDKRDCEYITLLERRSEGLACCTNARHPSCWNLVNGTV
VPLGEMRYAPSPDNLVLEGGDEVYTIKQYENGKIPRFRIRGSESELYTSDTV
MNQPIKATIHQDQAYDKIYFFREDNPNKPNLNVSRVQLCRDGGGESSL
SVSKWTFILKMLVCSDAATNKNFNLDVFLPPDQGMRTDRIYGVFSPNWNYSV
CYSGLDIDKVFSLKGYHSLSNPREFKCLPDQOPIPTETFOVADHRHPEVQORVE
PMGLPTLPHSKYHYQKVAHRMQASHGETFVLYLTDRGTIKHVVPGEQHSFA
FNMEIOPFRRAAIOTMSLDAERRKLYVSSOWEYSOVPLDLCEVYGGCHCLMSRD
PYCWDGRCISYSSERSVLOSINPAEPHKECPNPKPKAPLOKVS LAPNSRYLSC
PYESRHATYSWRKHNVEQSCPEGHGSPNCILFIENLTAAQIGHIFCEAQEGSYFNEA
QHWQLLPEDGIMAEHLHGHACALAAASLWLGVLPTLTGLLVH"
18..155
sig_peptide
misc_feature
156..212
/notes="Region: N-terminal peptide"
156..196i
/product="GPI-anchored membrane protein CDw108"
330..341
/notes="glycosylation site"
/evidence=not_experimental
486..497
/notes="glycosylation site"
/evidence=not_experimental
789..800
/notes="glycosylation site"
/evidence=not_experimental
816..824
/notes="Region: RGD cell attachment sequence"
1005..1016
/notes="glycosylation site"
/evidence=not_experimental
1821..1832
/notes="glycosylation site"
/evidence=not_experimental
1959..2015
/notes="Region: GPI-anchor motif"
BASE COUNT 522 a 850 c 769 g 520 t
ORIGIN
```

Query Match 100.0%; Score 1902; DB 12; Length 2661;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 CTGCTGCTGCTCTCGGCGCGCGCTCGCCCGGCGCACCTAAGGAGCGGACCC 60
Db 114 CTGCTGCTGCTCTCGGCGCGCGCTCGCCCGGCGCACCTAAGGAGCGGACCC 173
Qy 61 CGCATCTTCGCGCTCTGGAAGGCCATGTAGGCGAGGACCGGGTGGACTTTGGCCAGACT 120
Db 174 CGCATCTTCGCGCTCTGGAAGGCCATGTAGGCGAGGACCGGGTGGACTTTGGCCAGACT 233
Qy 121 GAGCCGACACAGGTGCTTTTCCAGGACCGAGGAGCTCTCTGTGTGGGTGGGAGGACGT 180
Db 234 GAGCCGACACAGGTGCTTTTCCAGGACCGAGGAGCTCTCTGTGTGGGTGGGAGGACGT 293
Qy 181 GCAAGTCTACCTCTTTGACTTCCCGGAGGCAAGACCATCTGTGCCACGGTGAAT 240
Db 294 GCAAGTCTACCTCTTTGACTTCCCGGAGGCAAGACCATCTGTGCCACGGTGAAT 333
Qy 241 ATCGGCTCCACAAAGGGTCTCTGTGATAAGCGGAGCTGCCAGAACTACACTCTC 300
Db 354 ATCGGCTCCACAAAGGGTCTCTGTGATAAGCGGAGCTGCCAGAACTACACTCTC 413
```

```
Qy 301 CTGGAGAGCGGAGTGAGGGGCTGCTGGCTGTGGCACCACCGCCGCGCACCCAGCTGC 360
Db 414 CTGGAGAGCGGAGTGAGGGGCTGCTGGCTGTGGCACCACCGCCGCGCACCCAGCTGC 473
Qy 361 TGGAACTCGTGAATGGCACTGTGGTGCACCTTGGCGAGATGAGAGGCTACGCCCCCTTC 420
Db 474 TGGAACTCGTGAATGGCACTGTGGTGCACCTTGGCGAGATGAGAGGCTACGCCCCCTTC 533
Qy 421 AGCCCGAGCAGAACTCCCTGGTTCCTTGAAGGGGAGGAGGTGATTTCACCATCCGG 480
Db 534 AGCCCGAGCAGAACTCCCTGGTTCCTTGAAGGGGAGGAGGTGATTTCACCATCCGG 593
Qy 481 AAGCAGAATACAAATGGGAAGATCCCTCGGTTCGCGCGCATCCGGGGCGAGAGTGAGCTG 540
Db 594 AAGCAGAATACAAATGGGAAGATCCCTCGGTTCGCGCGCATCCGGGGCGAGAGTGAGCTG 653
Qy 541 TACACCACTGATCTGTCATGCAAGAACCCACACAGTTTCATAAAGCCACCATCTCGTGACCAA 600
Db 654 TACACCACTGATCTGTCATGCAAGAACCCACACAGTTTCATAAAGCCACCATCTCGTGACCAA 713
Qy 601 GACCAAGCTTACGATGACAAAGATCTACTTCTTCCGAGAGGACAAATCCTGACAAAGAT 660
Db 714 GACCAAGCTTACGATGACAAAGATCTACTTCTTCCGAGAGGACAAATCCTGACAAAGAT 773
Qy 661 CCGTAGGCTCCTCTCAATGTGTCCTGTCGCGCGCATGTCGAGGGGGGACCAAGGGTGGG 720
Db 774 CCGTAGGCTCCTCTCAATGTGTCCTGTCGCGCGCATGTCGAGGGGGGACCAAGGGTGGG 833
Qy 721 GAAAGTTCACTGTGCTGCTCAAGTGAACACTTTTCTGAAAGCCCATGCTGTTATCGAGT 780
Db 834 GAAAGTTCACTGTGCTGCTCAAGTGAACACTTTTCTGAAAGCCCATGCTGTTATCGAGT 893
Qy 781 GATGCTGCCACCAACAAAGAACTTCAACAGGCTGCAAGAGCTCTTCTGCTCCCTGACCCC 840
Db 894 GATGCTGCCACCAACAAAGAACTTCAACAGGCTGCAAGAGCTCTTCTGCTCCCTGACCCC 953
Qy 841 AGCGGCGAGTGAGGGGACACAGGGTCTATGGTGTGTTTCTCAACCCCTTGGAACTACTCA 900
Db 954 AGCGGCGAGTGAGGGGACACAGGGTCTATGGTGTGTTTCTCAACCCCTTGGAACTACTCA 1013
Qy 901 GCGTCTGTGTGTTATTCCTCGGTGACATTGCAAGGCTTCCGTACCTCCTCCTCAAG 960
Db 1014 GCGTCTGTGTGTTATTCCTCGGTGACATTGCAAGGCTTCCGTACCTCCTCCTCAAG 1073
Qy 961 GGTACCACTCAAGCCTTCCCAACCGCGGCTGGAAGTGGCTTCCAGACGACGAGCCG 1020
Db 1074 GGTACCACTCAAGCCTTCCCAACCGCGGCTGGAAGTGGCTTCCAGACGACGAGCCG 1133
Qy 1021 ATACCACAGAGACCTTCCAGGTGGCTGACCTCACCCAGAGGTGGCGAGAGGGTGGAG 1080
Db 1134 ATACCACAGAGACCTTCCAGGTGGCTGACCTCACCCAGAGGTGGCGAGAGGGTGGAG 1193
Qy 1081 CCCATGGGCGCTCTGAAGAGCGCATTTGCTTCCACTCTAAATACCCTACCAGAAAGTGGCC 1140
Db 1194 CCCATGGGCGCTCTGAAGAGCGCATTTGCTTCCACTCTAAATACCCTACCAGAAAGTGGCC 1253
Qy 1141 GTCCACCGCATGCAAGCCACGCCGAGAGACTTTTCATGTGCTTTTACCTAACTACAGAC 1200
Db 1254 GTCCACCGCATGCAAGCCACGCCGAGAGACTTTTCATGTGCTTTTACCTAACTACAGAC 1313
Qy 1201 AGGGGCACTATCCACAAGGTGGTGGAAACCGGGGAGCAGGACAGCTTCGCTTCAAC 1260
Db 1314 AGGGGCACTATCCACAAGGTGGTGGAAACCGGGGAGCAGGACAGCTTCGCTTCAAC 1373
Qy 1261 ATCATGGAGATCCAGCCCTTCCCGCGCGGTGCCATCCAGACCATGTGCTGTGATGCT 1320
Db 1374 ATCATGGAGATCCAGCCCTTCCCGCGCGGTGCCATCCAGACCATGTGCTGTGATGCT 1433
Qy 1321 GAGCGGAGGAAGCTGTATGTAGCTCCCGAGTGGAGGTGAGCCAGGTGCCCTTGGACCTG 1380
Db 1434 GAGCGGAGGAAGCTGTATGTAGCTCCCGAGTGGAGGTGAGCCAGGTGCCCTTGGACCTG 1493
Qy 1381 TGTGAGGTCTATGGCGGGGCTGCCAGGTTGCCTCATGTCCCGGAGACCCCTACTCGGCG 1440
```

```
|||||
Db 1494 TGTGAGGCTTATGGGGGGTGCACAGTTCCTCATGTCCGAGACCCCTACTCGGC 1553
Qy 1441 TGGACCAAGCCGCTGATCTCATCTACAGTCCCGAACGGTGAAGTCTGCAATCCATT 1500
Db 1554 TGGACCAAGCCGCTGATCTCATCTACAGTCCCGAACGGTGAAGTCTGCAATCCATT 1613
Qy 1501 AATCAGCGGAGCCACACAGAGGTGTCCCAACCCCAACACAGACAGGCCCCACTGCAG 1560
Db 1614 AATCAGCGGAGCCACACAGAGGTGTCCCAACCCCAACACAGACAGGCCCCACTGCAG 1673
Qy 1561 AAGTTTCCCTGGCCCAAACTCTCGTACTACCTGAGTCCCGCCATGGAATCCCGCCAC 1620
Db 1674 AAGTTTCCCTGGCCCAAACTCTCGTACTACCTGAGTCCCGCCATGGAATCCCGCCAC 1733
Qy 1621 GCCACCTACTATGGCGCCACACAGAGACGTGAGACAGAGCTGCGAACCTGGTCCACAG 1680
Db 1734 GCCACCTACTATGGCGCCACACAGAGACGTGAGACAGAGCTGCGAACCTGGTCCACAG 1793
Qy 1681 AGCCCCAACTGCATCTCTTTCATCGAGAACCTCAGCGCGACAGTACGGCCACTACTTC 1740
Db 1794 AGCCCCAACTGCATCTCTTTCATCGAGAACCTCAGCGCGACAGTACGGCCACTACTTC 1853
Qy 1741 TGGAGGCCAGGAGGCTCTCTACTTCCGAGAGCTCAGCAGTGGCAGCTGCTGCCCGAG 1800
Db 1854 TGGAGGCCAGGAGGCTCTCTACTTCCGCGAGGCTCAGCAGTGGCAGCTGCTGCCCGAG 1913
Qy 1801 GACGGCATATGGCGGAGCAGCTGCTGGGTATGCTGTCGCTGGCGGCTCCCTCTGG 1860
Db 1914 GACGGCATATGGCGGAGCAGCTGCTGGGTATGCTGTCGCTGGCGGCTCCCTCTGG 1973
Qy 1861 CTGGGGTGTGTCGCCACACACTCCTTGGCTTGGCTGCTGCTCCAC 1902
Db 1974 CTGGGGTGTGTCGCCACACTCCTTGGCTTGGCTGCTGCTCCAC 2015

RESULT 2
LOCUS AF071542 2594 bp mRNA PRI 01-DEC-1998
DEFINITION Homo sapiens semaphorin K1 mRNA, complete cds.
ACCESSION AF071542
NID g3930578
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2594)
AUTHORS Xu, X., Ng, S., Wu, Z. L., Nguyen, D., Homburger, S., Seidel-Dugan, C.,
Ebens, A., and Luo, Y.
Homo semaphorin K1 is glycosylphosphatidylinositol-linked and
defines a new subfamily of viral-related semaphorins
J. Biol. Chem. 273 (35), 22428-22434 (1998)
98380763
JOURNAL 2 (bases 1 to 2594)
MEDLINE Xu, X. and Luo, Y.
REFERENCE Direct Submission
AUTHORS Xu, X. and Luo, Y.
JOURNAL Submitted (10-JUN-1998) Biology, Exelixis Pharmaceuticals, Inc., 260
Littlefield Ave., South San Francisco, CA 94080, USA
REFERENCE 3 (bases 1 to 2594)
AUTHORS Xu, X. and Luo, Y.
JOURNAL Direct Submission
TITLE Submitted (30-NOV-1998) Biology, Exelixis Pharmaceuticals, Inc., 260
Littlefield Ave., South San Francisco, CA 94080, USA
REMARK Sequence update by submitter
COMMENT On Dec 1, 1998 this sequence version replaced gi:3493630.
FEATURES Location/Qualifiers
source 1..2594
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q22"
1..2001
CDS
```

```
/note="sema K1; glycosylphosphatidylinositol-linked
protein; similar to Alcelaphine herpesvirus 1 semaphorin"
/codon_start=1
/product="semaphorin K1"
/db_xref="PID:g3930579"
/translation="MTPPPGRAAPSAPRAPVGPPLRLGLPLRLRLLLMAAASA
OGHLRSGPRIFAVKGVQDRVDFGTEPTHTVLFHEPSSSVVWGRGKVLFDPE
KNASVTVNIGSTKGLKDCENITLLERRSGLLACGNTNARHPSCNWLVNGTV
VPIGEMRGVAFPSDLSVLFEGDEVISTIRKOEYNGKIPRRIRGSELYSDTV
MNPQRTIKATIVHQDQYDKIYFFREDNPDKNPEAPLNVSRVQLCRDQGESL
SVKNTFLKMLVQDAAATKNFRLQVLLPDPGOWRDRVTVGVFSPNPNY SAV
CVSLGIDIKVFTSSLKPNRPGKCLPDDOPIPTETFOVADREPEVAORVE
PMGLTPLFHSKYHVQKAVHBMQASHGETFHVLYLTDRGTIHKVVGGEHSEFA
FNIMETQPRRAAIIOTMSLDAERRKLYVSSQWESVQVPLDLCEVVGCGGLMSRD
PYGWDQGRICISIISSERSVLSINPAEPHKECPNPKPQKLVSLAPNSRYLSC
PMESRATYSWRHKNVEQSEFQSPNCIEIENLTAQYGHYFCEAEGSYFREA
OHQWLLPEDGIMAEHLHGACLAASLWLVPLTLGLLVH"
BASE COUNT 514 a 814 c 756 g 510 t
ORIGIN

Query Match 100.0%; Score 1902; DB 12; Length 2594;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGCTGCTGCTGCTGCGGGCGCGCCCTCCGCCAGGCGCCACCTAAGAGCGGACCC 60
Db 97 CTGCTGCTGCTGCTGCGGGCGCGCCCTCCGCCAGGCGCCACCTAAGAGCGGACCC 156
Qy 61 CGCATCTTCGCCGCTCGAAGGCCATGTAGGGCAGACCGGGTGACCTTTGGCCAGACT 120
Db 157 CGCATCTTCGCCGCTCGAAGGCCATGTAGGGCAGACCGGGTGACCTTTGGCCAGACT 216
Qy 121 GAGCCGACACGGTGTCTTTCCAGGAGCGAGCTCTCTGTGTGGTGGGAGGACGT 180
Db 217 GAGCCGACACGGTGTCTTTCCAGGAGCGAGCTCTCTGTGTGGTGGGAGGACGT 276
Qy 181 GGCNAGTCTACCTCTTTGACTTCCCGAGGGCAAGACGATCTGTGCGCACGGTGAAT 240
Db 277 GGCNAGTCTACCTCTTTGACTTCCCGAGGGCAAGACGATCTGTGCGCACGGTGAAT 336
Qy 241 ATGGCTCCACAAAGGGTCTCTGTCTGATTAAGGGGACCTCCGAGAACTACATCTCTC 300
Db 337 ATGGCTCCACAAAGGGTCTCTGTCTGATTAAGGGGACCTCCGAGAACTACATCTCTC 396
Qy 301 CTGAGAGCGGAGTGAAGGGCTGCTGCTGTGGCAACCAACGCCGCCAGCTGC 360
Db 397 CTGAGAGCGGAGTGAAGGGCTGCTGCTGTGGCAACCAACGCCGCCAGCTGC 456
Qy 361 TGGNACCTGTGAATGGCAGTGTGGTCCACTTGGCGAGATGAGAGGCTACGCCCTTC 420
Db 457 TGGNACCTGTGAATGGCAGTGTGGTCCACTTGGCGAGATGAGAGGCTACGCCCTTC 516
Qy 421 AGCCCGGACGAGACTCCCTGGTCTGTTTGAAGGGGACGAGGTGTATTCACCATCCGG 480
Db 517 AGCCCGGACGAGACTCCCTGGTCTGTTTGAAGGGGACGAGGTGTATTCACCATCCGG 576
Qy 481 AAGCAGGAATACAAATGGGAAGATCCCTCGGTTCGCCCGCATCCGGGGCGAGGTGAGCTG 540
Db 577 AAGCAGGAATACAAATGGGAAGATCCCTCGGTTCGCCCGCATCCGGGGCGAGGTGAGCTG 636
Qy 541 TACACCACTGATCTGTCTGATGCAAGCCACAGTTCTATCAAGCCACCATCTGTCACCAA 600
Db 637 TACACCACTGATCTGTCTGATGCAAGCCACAGTTCTATCAAGCCACCATCTGTCACCAA 696
Qy 601 GACCAGCTTACGATGACAGATCTACTACTTCTTCGAGAGGACAACTCTGCAAGAAT 660
Db 697 GACCAGCTTACGATGACAGATCTACTACTTCTTCGAGAGGACAACTCTGCAAGAAT 756
Qy 661 CCTGAGGCTCCTCTCAATGTGCCGTGTGCCAGTTGTGCGAGGGGACCGAGGTGGG 720
Db 757 CCTGAGGCTCCTCTCAATGTGCCGTGTGCCAGTTGTGCGAGGGGACCGAGGTGGG 816
```

Qy	721	GAAGTTCACTGTCAGTCTCCAAAGTGGAAACACTTTTCTGAAAGCCATGCTGGTATGCAGT	780
Db	817	GAAGTTCACTGTCAGTCTCCAAAGTGGAAACACTTTTCTGAAAGCCATGCTGGTATGCAGT	876
Qy	781	GATGCTGCCACCAACAAAGAACTTCAACAGGCTGCAAGACGTCTTCTTGGCTCCCTGACCC	840
Db	877	GATGCTGCCACCAACAAAGAACTTCAACAGGCTGCAAGACGTCTTCTTGGCTCCCTGACCC	936
Qy	841	AGGGCCAGTGGAGGACACCAAGGCTCTATGCTGTTTCTTCCAACCCCTGGAACACTACTCA	900
Db	937	AGGGCCAGTGGAGGACACCAAGGCTCTATGCTGTTTCTTCCAACCCCTGGAACACTACTCA	996
Qy	901	GCGCTGTGTGTTATTCCTCTGGTGACATTGACAAGGTCTTCCGTACTCTCTCACTCAAG	960
Db	997	GCGCTGTGTGTTATTCCTCTGGTGACATTGACAAGGTCTTCCGTACTCTCTCACTCAAG	1056
Qy	961	GGCTTACCACCTCAAGCCTTCCCAACCCGGCGCTGCAAGTGCCCTCCAGACACGACCG	1020
Db	1057	GGCTTACCACCTCAAGCCTTCCCAACCCGGCGCTGCAAGTGCCCTCCAGACACGACCG	1116
Qy	1021	ATACCCACAGAGACTTCCAGGTGGCTGACCTCACCCAGAGGTGGCGACAGGGTGGAG	1080
Db	1117	ATACCCACAGAGACTTCCAGGTGGCTGACCTCACCCAGAGGTGGCGACAGGGTGGAG	1176
Qy	1081	CCGATGGGGCTCTGAAGACGCCATTGTTTCCACTCTAAATACCACTACCAAGAAAGTGCC	1140
Db	1177	CCGATGGGGCTCTGAAGACGCCATTGTTTCCACTCTAAATACCACTACCAAGAAAGTGCC	1236
Qy	1141	GTCCACCGCATGCAAGCCAGCCACGGGAGACCTTTTCATGTGCTTTACCTTAACCTACAGAC	1200
Db	1237	GTCCACCGCATGCAAGCCAGCCACGGGAGACCTTTTCATGTGCTTTACCTTAACCTACAGAC	1296
Qy	1201	AGGGCACTATCCACAAGTGTGGAAACCGGGAGCAGGACACAGCTTCGCCCTTCAAC	1260
Db	1297	AGGGCACTATCCACAAGTGTGGAAACCGGGAGCAGGACACAGCTTCGCCCTTCAAC	1356
Qy	1261	ATCATGGAGATCCAGCCCTTCGCCCGCGCGGTGCCATCCAGACCATGTGCTGTGGATGCT	1320
Db	1357	ATCATGGAGATCCAGCCCTTCGCCCGCGCGGTGCCATCCAGACCATGTGCTGTGGATGCT	1416
Qy	1321	GAGCGGAGGAGCTGTATGTAGCTCTCCAGTGGGAGGTGAGCCAGGTGCCCCCTGGACCTG	1380
Db	1417	GAGCGGAGGAGCTGTATGTAGCTCTCCAGTGGGAGGTGAGCCAGGTGCCCCCTGGACCTG	1476
Qy	1381	TGTGAGTCTATGGCGGGGCTGCCACGGTTCGCTCATGTCCCGAGACCCCTACTTGGCG	1440
Db	1477	TGTGAGTCTATGGCGGGGCTGCCACGGTTCGCTCATGTCCCGAGACCCCTACTTGGCG	1536
Qy	1441	TGGGACCAAGCGCGCTGCATCTCCATCTACAGCTCCGAACGGTTCAGTGTCAATATCCATT	1500
Db	1537	TGGGACCAAGCGCGCTGCATCTCCATCTACAGCTCCGAAGGGTTCAGTGTCAATATCCATT	1596
Qy	1501	AATCCAGCCAGCCACACAAGGAGTGTCCCAACCCCAACACAGACAGGCCCCACTGCAG	1560
Db	1597	AATCCAGCCAGCCACACAAGGAGTGTCCCAACCCCAACACAGACAGGCCCCACTGCAG	1656
Qy	1561	AAGTTTTCCCTGGCCCCAAACTCTCGCTACTACTTACCTGAGTGCCTCATGGAATCCGGCAC	1620
Db	1657	AAGTTTTCCCTGGCCCCAAACTCTCGCTACTACTTACCTGAGTGCCTCATGGAATCCGGCAC	1716
Qy	1621	GCCACCTACTCATGGGCGCCACAAGAGAACTGTGGAGCAGAGCTGCAACCTGGTCAACAG	1680
Db	1717	GCCACCTACTCATGGGCGCCACAAGGAGAACTGTGGAGCAGAGCTGCAACCTGGTCAACAG	1776
Qy	1681	AGCCCAACTGCATCTGTTTCATCGAGAACTTCACGGGGCAGCAGTAGCGCCACTACTTC	1740
Db	1777	AGCCCAACTGCATCTGTTTCATCGAGAACTTCACGGGGCAGCAGTAGCGCCACTACTTC	1836
Qy	1741	TGGAGGCCACAGAGGCTCTACTTCCGAGGCTCAGCACTGGCAGCTGCTGCCCGAG	1800
Db	1837	TGGAGGCCACAGAGGCTCTACTTCCGAGGCTCAGCACTGGCAGCTGCTGCCCGAG	1896
Qy	1801	GAGGGCATCATGCGCCAGCAGCCTGTGGGTCTATGCTGTGCCCTGGCGCCTCCCTCTGG	1860

Db	1897	GACGSCATCATGGCGGACACCTGCTGGGTTCATGCTGGCTGGCGGCTCCCTCTGG	1956
Qy	1861	CTGGGGTGGTGGCCACACTCCTCTTGGCTTGGTGGTCCAC	1902
Db	1957	CTGGGGTGGTGGCCACACTCCTCTTGGCTTGGTGGTCCAC	1998
RESULT	3		
LOCUS	AF069493	2661 bp mRNA PRI	03-DEC-1998
DEFINITION	Homo sapiens GPI-anchored membrane protein CDw108 precursor, mRNA, complete cds.		
ACCESSION	AF069493		
NID	g3947939		
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 2661)		
AUTHORS	Yamada, A., Kubo, K., Takeshita, T., Harashima, N., Kawano, K., Sagawa, K., Suganura, K. and Itoh, K.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUN-1998) Immunology, Kurume University School of Medicine, Asahi-machi 67, Kurume, Fukuoka 830, Japan		
FEATURES	Location/Qualifiers		
source	1..2661		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="15"		
	/map="between WI-6247 and CHLC.GATA85D02"		
	/cell_line="HPB-ALL"		
	/cell_type="T lymphocyte"		
CDS	18..2018		
	/note="putative adhesion molecule; lymphocyte activation antigen"		
	/codon_start=1		
	/product="GPI-anchored membrane protein CDw108 precursor"		
	/db_xref="PID:g3947940"		
	/translation="MTPTPPGRAPSAPRAVPGPPARLGLPLRLRLLLLLWAAAAASRQGLRSGRPIFAWKHGVDGFGOTEPHTVLFPFGSSVWVGKRGKYLDFDFEFGKNASVTVAIGSTKSCLDKRCENYITLLRRSEGLACGTNARHPSCWNLVNGTVPLGEMRGYAPFSDENSLVDFEGDEVYITRQYNGKIPFRFRIREGSELYTSDTVMQNQFTKATIVHQDOAYDDKIYVFREDNPDKNEAPLNVSRVAQLCRGDQGGSLLSVSKNQFTFLKAMLVQSDAAATKNFNRLQDYVLLDPDSGMWRDTRVGVFNSPNYSAVCVSLGDIDKVFRTSSLKGYSLSLPNRPCKLPDQOPIETETFOVADRPEVAQRVEPMGLKPTLPFHSYHKYQAVHMQASGHGTFHVLYLTDRGTIHKVVEPGEQEHSAFTNMEIQPFRAAIIQMSDADERRLYVSSQWVPLDLCVGGGGCGGCLMSFAPYCGDGRGISIYSSYRVSQISINPAEHPGCPNPKAPLQKVLAPNSRYLSLCPMSRHATYSWRHKNVEQSCPECHQSPNCILFIENLTAAQYGHYFCEAQEGSYFREAQHWQLPEDGIMAEHLGLHACALAAASLWGLVLTLTGLLVH"		
	18..155		
	156..212		
sig_peptide	156..212		
misc_feature	156..1961		
mat_peptide	/note="Region: N-terminal peptide"		
misc_feature	/product="GPI-anchored membrane protein CDw108"		
	330..341		
	/note="glycosylation site"		
misc_feature	/evidence=not_experimental		
	486..497		
	/note="glycosylation site"		
misc_feature	/evidence=not_experimental		
	789..800		
	/note="glycosylation site"		
misc_feature	/evidence=not_experimental		
	816..824		
	/note="Region: RGD cell attachment sequence"		
misc_feature	1005..1016		
misc_feature	/note="glycosylation site"		
	/evidence=not_experimental		
misc_feature	1821..1832		
	/note="glycosylation site"		

KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2594)
AUTHORS Xu, X., Ng, S., Wu, Z.L., Nguyen, D., Homburger, S., Seidel-Dugan, C.,
Ebens, A., and Luo, Y.
TITLE Human semaphorin K1 is glycosylphosphatidylinositol-linked and
defines a new subfamily of viral-related semaphorins
J. Biol. Chem. 273 (35), 22428-22434 (1998) --
JOURNAL 98380463
MEDLINE 2 (bases 1 to 2594)
REFERENCE Xu, X. and Luo, Y.
AUTHORS Direct Submission
TITLE Submitted (10-JUN-1998) Biology, Exelixis Pharmaceuticals, Inc., 260
JOURNAL Littlefield Ave, South San Francisco, CA 94080, USA
REFERENCE 3 (bases 1 to 2594)
AUTHORS Xu, X. and Luo, Y.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-1998) Biology, Exelixis Pharmaceuticals, Inc., 260
JOURNAL Littlefield Ave, South San Francisco, CA 94080, USA
REMARK Sequence update by submitter
COMMENT On Dec 1, 1998 this sequence version replaced gi:3493630.
FEATURES
source
1..2594
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="19q22"
1..2001
/note="sema K1; glycosylphosphatidylinositol-linked
protein; similar to Alcelaphine herpesvirus 1 semaphorin"
/codon_start=1
/product="semaphorin K1"
/db_xref="PID:g3930579"
/translation="MTPTPPGGAAPSAAPRVPVGPAPRLRLRLRLRLRLWAAASA
QGLHSGPRIFAVMKGVHQQDRVDFQTEPHVLPHEPGSSVWVGGRGKVLDFDPE
GKNSVRTVNIIGSTGSLDKRDCENYITLLERRSEGLLACTNARHPNSLVNVTG
VPLGEMRYAPSPDENSLVLEGEDEVSTIRKQYNGKIPFRIRIGSESLYISDTV
MNPQFIKAIQVQADYDKIYFFREDNPKNEAPLNVRVAQLKRGDGGSSLV
SVSKWNTFLKMLVCSDAATNKNFNRLQDVFLLPDPGQWRDTRYGVFSNPVYSV
CYSGLGDIKVFRTSLKGYHSLNPAPKLPDQPIPTETFOVADRHPEVAQRVE
PMGLKPLTFHSAKHYQKVAVHMOASHGETHVLVLTDRGTIKHVPGEQHSFA
FNIMETQPRRAAIQTMSLDAERKLVSOWMEYSQVPLDLCEVYGGSGCLMSRD
PCGWDQGRCSIIYSERSVLOSINPAEPHPCPNPKDPKAPLQKVSAPNSRYLSC
PMESHAITYSRHKNVQSCPEGHSPNCILFIENLTAQQIYGHYFCEAQEGSYFREA
QHWQLLPEDGINAEHLGHACALASLMGLVLPITLLGLLVH"
BASE COUNT 514 a 814 c 756 g 510 t
ORIGIN

Query Match 100.0%; Score 1902; DB 42; Length 2594;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTGCTGCTCTGGGGCGCGCGCTCCGCCAGGCGCACCTAAGGAGCGGACCC 60
|||||
DB 97 CTGCTGCTGCTCTGGGGCGCGCGCTCCGCCAGGCGCACCTAAGGAGCGGACCC 156
|||||
QY 61 CGCATCTTCGCGCTCTGGAAAGGCCATGTAGGCGCAGGACCGGCTTGCCAGACT 120
|||||
DB 157 CGCATCTTCGCGCTCTGGAAAGGCCATGTAGGCGCAGGACCGGCTTGCCAGACT 216
|||||
QY 121 GAGCCGACACGGTGTCTTTCCACGAGCCAGGAGCTCTCTCTGTGTGGTGGGAGGACGT 180
|||||
DB 217 GAGCCGACACGGTGTCTTTCCACGAGCCAGGAGCTCTCTCTGTGTGGTGGGAGGACGT 276
|||||
QY 181 GCGAGGTCTACCTCTTTGACTTCCCGAGGGGCAAGACGATCTGTGGCAGGGTGAAT 240
|||||
DB 277 GCGAAGGTCTACCTCTTTGACTTCCCGAGGGGCAAGACGATCTGTGGCAGGGTGAAT 336
|||||
QY 241 ATCGGCTCCACAAAGGGTCTGTCTGGATAAGCGGAGCTGCGAGAACTACATCACTCTC 300
|||||

Qy	721	GAAGTTCACTGTCAGTCTCCAAAGTGGAAACACTTTTCTGAAAGCCATGCTGGTATGCAGT	780
Db	829	GAAGTTCACTGTCAGTCTCCAAAGTGGAAACACTTTTCTGAAAGCCATGCTGGTATGCAGT	888
Qy	781	GATGCTGCCACCAACAAGAACITTCACAGGCTGCAAGACGTTCTTCCTGCTCCCTGACCCC	840
Db	889	GATGCTGCCACCAACAAGAACITTCACAGGCTGCAAGACGTTCTTCCTGCTCCCTGACCCC	948
Qy	841	AGCGGCCAGTGGAGGACACACAGGCTCATGTGTGTTTCTCCAAACCCTCGAACTACTCA	900
Db	949	AGCGGCCAGTGGAGGACACACAGGCTCATGTGTGTTTCTCCAAACCCTCGAACTACTCA	1008
Qy	901	GCGGTCTGTGTATTTCCCTCGGTGACATTGACAAGGTCTTTCCGTACTCTCTCACCTCAAG	960
Db	1009	GCGGTCTGTGTATTTCCCTCGGTGACATTGACAAGGTCTTTCCGTACTCTCTCACCTCAAG	1068
Qy	961	GGCTACCACCTCAAGCCTTCCCAACCGCGGCTGCAAGTGCCTCCAGACACAGCAGCCG	1020
Db	1069	GGCTACCACCTCAAGCCTTCCCAACCGCGGCTGCAAGTGCCTCCAGACACAGCAGCCG	1128
Qy	1021	ATACCCACAGAGACCTTCCAGGTGGCTGACCGTCAACCAGAGGTGGCGCAGAGGGTGGAG	1080
Db	1129	ATACCCACAGAGACCTTCCAGGTGGCTGACCGTCAACCAGAGGTGGCGCAGAGGGTGGAG	1188
Qy	1081	CCCATGGGGCTCTGAAGACGCCATTGTTTCCACTCTAAATACCACCTACCAGAAGTGGCC	1140
Db	1189	CCCATGGGGCTCTGAAGACGCCATTGTTTCCACTCTAAATACCACCTACCAGAAGTGGCC	1248
Qy	1141	GTCCACCGCATGCAAGCCAGCCACGGGAGACCTTTCAATGTGCTTTACCTAACTACAGAC	1200
Db	1249	GTTCACCGCATGCAAGCCAGCCACGGGAGACCTTTCAATGTGCTTTACCTAACTACAGAC	1308
Qy	1201	AGGGCACAATACTCACAAAGTGTGTGAACCGGGGAGCAGGAGCACAGCTTCGGCTTCAAC	1260
Db	1309	AGGGCACAATACTCACAAAGTGTGTGAACCGGGGAGCAGGAGCACAGCTTCGGCTTCAAC	1368
Qy	1261	ATCATGAGATCCAGCCCTTCCGCGCGCGGTGCCATCCAGACCATGTGCGCTGGATGCT	1320
Db	1369	ATCATGAGATCCAGCCCTTCCGCGCGCGGTGCCATCCAGACCATGTGCGCTGGATGCT	1428
Qy	1321	GAGCGGAGGAAGCTGTATGTGAGCTCCCAAGTGGGAGGTGACCCAGGTGCCCTTGGACCTG	1380
Db	1429	GAGCGGAGGAAGCTGTATGTGAGCTCCCAAGTGGGAGGTGACCCAGGTGCCCTTGGACCTG	1488
Qy	1381	TGTGAGTCTATGCGGGGGCTGCCACGGTTGCGCTCATGTCCCGAGACCCCTACTCGGCG	1440
Db	1489	TGTGAGTCTATGCGGGGGCTGCCACGGTTGCGCTCATGTCCCGAGACCCCTACTCGGCG	1548
Qy	1441	TGGSACCAAGCGCGCTGCATCTCACAGCTCCGAACGGTCACTGCTGCAATCCATT	1500
Db	1549	TGGSACCAAGCGCGCTGCATCTCACAGCTCCGAACGGTCACTGCTGCAATCCATT	1608
Qy	1501	AATCCAGCCGAGCCACACAAGAGTGTCCCAACCCCAACACAGACAGGCCCACTGCAG	1560
Db	1609	AATCCAGCCGAGCCACACAAGAGTGTCCCAACCCCAACACAGACAGGCCCACTGCAG	1668
Qy	1561	AAGTGTTCCTGCGCCCAACACTCTCGCTACTACTGAGCTGCGCCCATGGAATCCGCCAC	1620
Db	1669	AAGTGTTCCTGCGCCCAACACTCTCGCTACTACTGAGCTGCGCCCATGGAATCCGCCAC	1728
Qy	1621	GCCACCTACTCATGGCGCCACAAAGAGACGTGGAGCAGAGCTGCGAACCTTGGTCAACAG	1680
Db	1729	GCCACCTACTCATGGCGCCACAAAGAGACGTGGAGCAGAGCTGCGAACCTTGGTCAACAG	1788
Qy	1681	AGCCCCAATGCATCTGTTTCAATCGAGAACCTCTACGGCGCAGCAGTACGGCCACTACTTC	1740
Db	1789	AGCCCCAATGCATCTGTTTCAATCGAGAACCTCTACGGCGCAGCAGTACGGCCACTACTTC	1848
Qy	1741	TGGGAGGCCCCAGAGGGCTCTCTACTTCCCGAGGCTCAGCACTGGCAGCTGCTGCCCGAG	1800
Db	1849	TGGGAGGCCCCAGAGGGCTCTCTACTTCCCGAGGCTCAGCACTGGCAGCTGCTGCCCGAG	1908
Qy	1801	GACGGCATATGCCCCAGCACCCTGCTGGGTATGCTGTGCGCTTGGCGGCGCTCCCTCTGG	1860

D	b	1909	GACGGCATATGCGCGAGCACCTGCTGGGTATGCCTTGCCCTGGCTGCCCTCCCTCTGG	1968
Q	y	1861	CTGGGGTGTGCCACACACTCACTTTGGCTTGCTGGTCACC	1902
D	b	1969	CTGGGGTGTGCCACACACTCACTTTGGCTTGCTGGTCACC	2010
R	E	S	U	L
RESULT	6			
AF030698				
LOCUS	AF030698	2636 bp	mRNA	PRI 04-SEP-1998
DEFINITION	Homo sapiens semaphorin L (SEMAL) mRNA, complete cds.			
ACCESSION	AF030698			
NID	g3523114			
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;			
REFERENCE	Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 2636)			
TITLE	Lange,C., Liehr,T., Goen,M., Gebhart,E., Fleckenstein,B. and Esser,A.			
JOURNAL	New eukaryotic semaphorins with close homology to semaphorins of DNA viruses			
MEDLINE	Genomics 51 (3), 340-350 (1998)			
REFERENCE	2 (bases 1 to 2636)			
AUTHORS	Esser,A., Liehr,T., Lange,C., Goen,M., Gebhart,E. and Fleckenstein,B.			
TITLE	Direct submission			
JOURNAL	Submitted (21-OCT-1997) Institut fuer Klinische und Molekulare Virologie, Friedrich Alexander Universitaet, Schlossgarten 4, Erlangen, Bavaria 91054, Germany			
FEATURES	Location/Qualifiers			
source	1..2636			
gene	/organism="Homo sapiens"			
CDS	/db_xref="taxon:9606"			
	/map="15p22.3-p23"			
	/chromosome="15"			
	1..2636			
	/gene="SEMAL"			
	13..2013			
	/gene="SEMAL"			
	/note="H-Sema-L"			
	/codon_start=1			
	/product="semaphorin L"			
	/db_xref="PID:g3523115"			
	/translation="MTPTPPPGRAAPSAPRVPGPARGLPRLRLRLLLWAAASA QQLRSGRPIFAVKGHVGDVRFGTEPHTVLFPHEPGSSVMVGKGVLFDFPE KGNASVTNIGTSGLCDLRDENEITLLERRSEGLACGTNAHPSCWNLVNSTV VPLNEWRVYAPISPDENSLVFEGDEVSTIRKOEYNGKIIPRRITRGESLYLTDIV MONPOLIKATIVHODAYDDKYIFREDNPKNPEAPLNVSVAQLCRGDGGRESSL SVSKGDIKAVMLVCSDAATNKFNKLQDFLLDPDGOWDRTRTVGTGVSFNPWNYSV FIVEGLDIDFKVFTSSLGYHSLSNPREGKCLDPOOQIPTETFOVADRHVEVAQRFE PMGLPTPLFSKHYYKVAVHRMQASHGETFHVLYLTLDRTGTHKXVPEPGQEHSIFA FMELIQTPFRRAAIQTMSLDARERKLYVSSOWEVSOPDLCLCEVYGSGCHGCLMSRD PMSGWDQCISITYSSERSVLOSINAEPHKECPNPKDKAPLOKVSILAPNSRYLYLSC PYCSRATTYSRKHNKEVSPGHQSPNCILFIENLTAQQVGHYFCEAQEGSYFREIA QHWQLLPDGIWAHELHLGHACALAASLWLGVLTPLTLGLLVH"			
BASE COUNT	536 a 835 c 764 g 501 t			
ORIGIN				
Query Match	99.7%	Score 1897.2:	DB 42:	Length 2636;
Best Local Similarity	99.8%:	Pred. No. 0:		
Matches 1899:	Conservative	0: Mismatches	3: Indels	0: Gaps 0:
Qy	1	CTGCTGCTGCTGCTCGGCGCGCGCCTCGCCCGAGGGCCACCTTAAGGACGGACCC	60	
D	b	109	CTGCTGCTGCTGCTGCGGCGCGCGCCTCGCCCGAGGGCCACCTTAAGGACGGACCC	168
Qy	61	CGCATCTTCGCGCTCTGGAAGGCCATGTAGGCAGGACCGGGTGGACTTTGGCCAGACT	120	

Db 169 CGCATCTTCGCGCTCTGGAAGGCCATGTAGGCGAGGACCGGGTGGACTTTTGGCCAGACT 228
QY 121 GAGCCGCACACGGTCTGTTTCCACAGCCAGCAGCTCTCTGTGTGGTGGGAGGAGT 180
Db 229 GAGCCGCACACGGTCTGTTTCCACAGCCAGCAGCTCTCTGTGTGGTGGGAGGAGT 288
QY 181 GGCAAGGTCTACCTCTTTGACTTCCCGGAGGCAAGACGATCTGTGCGACGGTGAAT 240
Db 289 GGCAAGGTCTACCTCTTTGACTTCCCGGAGGCAAGACGATCTGTGCGACGGTGAAT 348
QY 241 ATCGGCTCCACAAAGGGTCTCTGTCTGGATAAGCGGGACTCGGAGAACTACATCACTCTC 300
Db 349 ATCGGCTCCACAAAGGGTCTCTGTCTGGATAAGCGGGACTCGGAGAACTACATCACTCTC 408
QY 301 CTGGAGAGCGGAGTGAAGGGTCTGTGCGCTGTGCGACCAAGCGCCCGGACCCCAAGTGC 360
Db 409 CTGGAGAGCGGAGTGAAGGGTCTGTGCGCTGTGCGACCAAGCGCCCGGACCCCAAGTGC 468
QY 361 TGGAACTTGGTGAATGGACCTGTGTGTCACCTTGGCGAGATGAGAGGCTACGCCCTTC 420
Db 489 TGGAACTTGGTGAATGGACCTGTGTGTCACCTTGGCGAGATGAGAGGCTACGCCCTTC 528
QY 421 AGCCCGGAGGAGAACTCCCTGTTCTGTTTGAAGGGGACGAGGTGTATTCCACCATCGG 480
Db 529 AGCCCGGAGGAGAACTCCCTGTTCTGTTTGAAGGGGACGAGGTGTATTCCACCATCGG 588
QY 481 AAGCAGGAATACAAATGGGAAGATCCCTCGTTCGCGCGATCCGCGGGGAGAGTGAAGTG 540
Db 589 AAGCAGGAATACAAATGGGAAGATCCCTCGTTCGCGCGATCCGCGGGGAGAGTGAAGTG 648
QY 541 TACACAGTGATACCTGATCAGAACCCACAGTTCATCAAGCCACCATCGTGACCAAA 600
Db 649 TACACAGTGATACCTGATCAGAACCCACAGTTCATCAAGCCACCATCGTGACCAAA 708
QY 601 GACCAAGGTTACGATGACAAAGATCTACTACTTCTCCGAGAGGACAACTCTGACAAAGAT 660
Db 709 GACCAAGGTTACGATGACAAAGATCTACTACTTCTCCGAGAGGACAACTCTGACAAAGAT 768
QY 661 CTGAGGCTCTCTCAATGTGTCCGCTGTGGCCAGTGTGTGGAGGGGAGACAGGTTGG 720
Db 769 CTGAGGCTCTCTCAATGTGTCCGCTGTGGCCAGTGTGTGGAGGGGAGACAGGTTGG 828
QY 721 GAAAGTTACATCTGATCTCCAAAGTGGACACTTTTCTGAAAGCCATGCTGGTATGCAAT 780
Db 829 GAAAGTTACATCTGATCTCCAAAGTGGACACTTTTCTGAAAGCCATGCTGGTATGCAAT 888
QY 781 GATGTCGACCAACAAAGAACTTCAACAGGCTGCAAGAGCTTCTCTGCTCCCTGACCC 840
Db 889 GATGTCGACCAACAAAGAACTTCAACAGGCTGCAAGAGCTTCTCTGCTCCCTGACCC 948
QY 841 AGCGGCCAGTGGAGGACACCAAGGCTCTATGGTGTGTTTCTCCAAACCCCTGGAACCTCA 900
Db 949 AGCGGCCAGTGGAGGACACCAAGGCTCTATGGTGTGTTTCTCCAAACCCCTGGAACCTCA 1008
QY 901 GCGCTCTGTGTATCTCCCTCGTGACATTTGACAGAGTCTCCGTTACCTCTCACTCAAG 960
Db 1009 GCGCTCTGTGTATCTCCCTCGTGACATTTGACAGAGTCTCCGTTACCTCTCACTCAAG 1068
QY 961 GCGTACCACTCAAGCTTCCCAACCCGCGCTGCAAGTGTCTCCAGACCAAGGAGCGG 1020
Db 1069 GCGTACCACTCAAGCTTCCCAACCCGCGCTGCAAGTGTCTCCAGACCAAGGAGCGG 1128
QY 1021 ATACCCACAGAGACCTTCCAGGTGCTGACCGTCAACCCAGAGGTTGGCGAGAGGTGGAG 1080
Db 1129 ATACCCACAGAGACCTTCCAGGTGCTGACCGTCAACCCAGAGGTTGGCGAGAGGTGGAG 1188
QY 1081 CCCATGGGCTCTGAAGACGCCATTTGTCACCTTAATACCATACAGAAAGTGGCC 1140
Db 1189 CCCATGGGCTCTGAAGACGCCATTTGTCACCTTAATACCATACAGAAAGTGGCC 1248
QY 1141 GTCCACCGCATCAAGACGACGCGGAGACCTTTGCTGCTTTACCTACTACAGAC 1200
Db 1249 GTTACCGCATCAAGACGACGCGGAGACCTTTGCTGCTTTTACCTACTACAGAC 1308

QY 1201 AGGGCACTATCCACAGGTGGTGAACCCGGGAGGAGGAGCAGAGCTTGCCTTCAAC 1260
Db 1309 AGGGCACTATCCACAGGTGGTGAACCCGGGAGGAGGAGCAGAGCTTGCCTTCAAC 1368
QY 1261 ATCATGGAGATCCAGCCCTTCCGCGCGCGCTGCGATCCAGACCATGCTGCTGATGCT 1320
Db 1369 ATCATGGAGATCCAGCCCTTCCGCGCGCGCTGCGATCCAGACCATGCTGCTGATGCT 1428
QY 1321 GAGCGGAGAAAGCTGTATGTGAGCTTCCAGTGGGAGGTGAGCCAGTGCCTTGGACCTG 1380
Db 1429 GAGCGGAGAAAGCTGTATGTGAGCTTCCAGTGGGAGGTGAGCCAGTGCCTTGGACCTG 1488
QY 1381 TGTGAGTCTATGGCGGGGCTGCGACAGTGTGCTCATGTCCGAGAGCCCTACTGCGGC 1440
Db 1489 TGTGAGTCTATGGCGGGGCTGCGACAGTGTGCTCATGTCCGAGAGCCCTACTGCGGC 1548
QY 1441 TGGGACCAAGCCGCTGCTCATCTACAGTCCGAAAGCTGAGTGCCTGCTGCAATCCATT 1500
Db 1549 TGGGACCAAGCCGCTGCTCATCTACAGTCCGAAAGCTGAGTGCCTGCTGCAATCCATT 1608
QY 1501 AATCCAGCGAGCCACACAGGAGTGTCCCAACCCCAACCAAGCCGCTGCTGCAATCCATT 1560
Db 1609 AATCCAGCGAGCCACACAGGAGTGTCCCAACCCCAACCAAGCCGCTGCTGCAATCCATT 1668
QY 1561 AAGTTTCCCTGGCCCAAACTCTGCTACTACCTGAGTGCCTCATGGAATCCGCGCAC 1620
Db 1669 AAGTTTCCCTGGCCCAAACTCTGCTACTACCTGAGTGCCTCATGGAATCCGCGCAC 1728
QY 1621 GCCACTACTATGGCGCCACAGGAGAACTGAGGAGAGCTGCGAGAGCTGCGAACCTGCTG 1680
Db 1729 GCCACTACTATGGCGCCACAGGAGAACTGAGGAGAGCTGCGAACCTGCTGCTGCTGCTG 1788
QY 1681 AGCCCCAACTGCTCTTTCATGAGAACTCAGCGCGAGAGTACGCGGAGTACGCTTCTG 1740
Db 1789 AGCCCCAACTGCTCTTTCATGAGAACTCAGCGCGAGAGTACGCGGAGTACGCTTCTG 1848
QY 1741 TGCAGGCCCCAGGAGGCTCTTCTTCCGAGAGGCTGAGAGTACGAGTCTGCTGCGCGAG 1800
Db 1849 TGCAGGCCCCAGGAGGCTCTTCTTCCGAGAGGCTGAGAGTACGAGTCTGCTGCGCGAG 1908
QY 1801 GACGCACTATGGCGGAGCAGCTGCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
Db 1909 GACGCACTATGGCGGAGCAGCTGCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1968
QY 1861 CTGGGGTGTCTGCGCACACTCACTCTTGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1902
Db 1969 CTGGGGTGTCTGCGCACACTCACTCTTGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2010

RESULT 7

AF030699 1192 bp mRNA ROD 04-SEP-1998
LOCUS Mus musculus semaphorin L (SemaL) mRNA, partial cds.
DEFINITION AF030699
ACCESSION AF030699
NID 93523116
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1192)
Lange, C., Liehr, T., Goen, M., Gebhart, E., Fleckenstein, B. and
Ensser, A.
TITLE New eukaryotic semaphorins with close homology to semaphorins of
DNA viruses
JOURNAL Genomics 31 (3), 340-350 (1998)
MEDLINE 98389619
REFERENCE 2 (bases 1 to 1192)
Ensser, A., Liehr, T., Lange, C., Goen, M., Gebhart, E. and
Fleckenstein, B.
AUTHORS Direct Submission
TITLE Institut fuer Klinische und Molekulare
JOURNAL Submitted (21-OCT-1997)

Virologie, Friedrich Alexander Universitaet, Schlossgarten 4, Erlangen, Bavaria 91054, Germany
Location/Qualifiers 1. .1192 /organism="Mus musculus" /db_xref="taxon:10090" /map="9A3.3-B" /chromosome="9" gene 1. .>1192 /gene="SemaL" /evidence=not_experimental CDS 13. >1192 /gene="SemaL" /note="M-Sema-L" /codon_start=1 /evidence=not_experimental /product="semaphorin L" /db_xref="PID:g3523117" /translation="MTPTPPGRAAPSAPRVLVSLPARFGLRLRLLLLVFWVAASA OQHSRGPRISAYMKQDHVDFOPEPHTVLFHPGSFSVMVGGRGVYHFNPFEGK STRTVNIGTSKQCKDKPCGNITLLERNGNLVCVTNARKPSKNILVNDSPVMS LGSMKVAPESPENSLIVLPGEDEVSTIRKOEYNGIKIPREFIRSESELVTDTYMG NPQFIKATIVHQDAVDKKIYYFEREDNPDKNEAPLNYSVAQLCRGDGGGBSSLSV SKWNTLKALVCSLRATNRNFLQDFLFLPDFSGMDRTIRVIGFSNPWTNSAVCV YLSGLDIDRVERTSSDKGYHMGLSNRPGMCLPKKPQIPTETIFQVADSHPEVAQRVEPM "
BASE COUNT 254 a 355 c 336 g 247 t
ORIGIN
Query Match 44.9%; Score 853.8; DB 13; Length 1192; Best Local Similarity 87.7%; Pred. No. 4e-163; Matches 959; Conservative 0; Mismatches 122; Indels 12; Gaps 2;
Qy 1 CTGCTGTCTGCTCTTGGCGGCCCGCCCTCCGCCAGGCCACCTTAAGAGCGGACC 60 Db 109 CTTCTGTGGTGTTCTTGGTGGCGCGCCCTCCGCCAAGGCCACTCGAGAGCGGACC 168 Qy 61 CGCATCTTCGCGCTCGAAAGGCATGTAGGCGAGACGGGTGGACTTTGGCCAGACT 120 Db 169 CGCATCTTCGCGCTCGAA-----AGGCGAGCACATGTGGACTTTAGCCAGCT 219 Qy 121 GAGCCGCACCGGTGCTTTTCCACGACGACGAGCTCTCTGTGTGGTGGGAGGACGT 180 Db 220 GAGCCACACCGGTGCTTTTCCATGAGCGGCGAGCTTCTCTGTGSGTGGTGGAGT 279 Qy 181 GGCAAGTCTACCTCTTTGACTTCCGAGGGGAAGAACGCATCTGTGCGCAGCGTGAAT 240 Db 280 GGCAAGTCTACCACTTCAACTTCCGAGGGGAAGAATGCCCTCTGTGCGCAGCGTGAAC 339 Qy 241 ATCGGCTCCAAAGGGTCTGTCTGGATAAGCGGACTCGAGAGACTACATCACTCTC 300 Db 340 ATCGGCTCCAAAGGGTCTGTCTGAGACAACAGGACTGTGGGAATTACATCACTCTT 399 Qy 301 CTGAGAGCGGGAGTAGGGGCTGCTGCCTGTGGCACCAACGCCGCCGCCAGCTGC 360 Db 400 CTAGAAGCGGGGTANTGGCTGCTGTCTGCGACCAATGCCGGAAGCCAGCTGC 459 Qy 361 TGGAACTGTGTAATGGCACTGTGG---TGCCACTTGGCGAGATGAGAGGCTACGCCCCC 417 Db 460 TGGAACTTGGTGAATGACAGTGTGGTGATGTCACTTGGTGAGATGAAAGGCTATGCCCCC 519 Qy 418 TTACGCCGGACGAGAACTCCCTGGTCTGTTTGAAGGGGACGAGGTGATTTCCACCATC 477 Db 520 TTACGCCGGATGAGAACTCCCTGGTCTGTTTTGAAGGAGATGAAGTGTACTCTACCATC 579 Qy 478 CGGAAGCAGGAATACAAATGGGAAGATCCCTCGGTTCCGCCCATCCGGGGGAGAGTGAG 537 Db 580 CGGAAGCAGGANATACACGGGAAGATCCCTCGGTTTCGACCCATTCGGGGCGAGGTAA 639 Qy 538 CTGTACACCAAGTACTGTGATCGAAGCCACAGTTCAAGAGCCACCATCGTGCAC 597 Db 640 CTGTACAAAGTATACAGTATCGAAGCCACAGTTCAAGAGCCACCATCGTGTGCAC 699

QY	598	CAAGACGAGCCTTACGATGACAAGATTACTACTTCTCCGAGAGAGACAATCCTGACAAG	657
Db	700	CAAGACCAGCCCTATGATGATGAAGATTACTACTTCTCCGAGAAGACAACCTTGACAAG	759
QY	658	AATCTGTAGGCTCCTCTCAATGTGTCCCCTGTGGGCCCATTTGTGCAGGGGGGACACAGGGT	717
Db	760	AACCCCGAGGCTCCTCTCAATGTGTCCCAGTAGCCACAGTTGTGCAGGGGGGACACAGGGT	819
QY	718	GGGAAAGTTTCAGTGTCTCCAGTCCAAGTGGAACACTTTTTCTGAAAGCCCATGCTGGTATGCG	777
Db	820	GGTGAGAGTTCTGTGTCTCTCCAGTGGGAACACCTTCTTGAAGCCCATGTTGGTCTGCG	879
QY	778	AGTGTGCTGCCACCAACACAGAACTTCAACAGGCTCAAGAGCTCTTCTGCTCCCTGAC	837
Db	880	AGCGATGACGCCACCAACAGGAACCTTCAATCGSCTCAAGATGTCTTCTGCTCCCTGAC	939
QY	838	CCNAGGCCAGTGGAGGACACACAGGCTCTATGGTGTCTTCTCCNACCCCTGGAACTAC	897
Db	940	CCNAGTGGCCAGTGGAGAGATACACAGGCTCTATGGCGTTTCTTCCAACCCCTGGAACTAC	999
QY	898	TCAGCGCTGTGTGTATTCCTTCGTCGTGACATTGACAAGTCTTCGGTACCTCCCTCACCTC	957
Db	1000	TCAGCTGTGTGTGTATTCGCTGTGTGACATTGACAGAGTCTTCGGTACCTCATCGCTC	1051
QY	958	AAGGGTACCACCTCAAGCCTTCCCAACCGCGGCTTGGCAAGTGGCTCCCGACACGACGAC	1011
Db	1060	ANAGGCTACCACATGGGCTTTCCACACCTCGACCTGGCATGTGCTCCCAAAAAGCAG	1111
QY	1018	CCGATACCCACAGAGACCTTCCAGTGGCTGACCGTFCACCCAGAGAGTGGCGCAGAGGGTG	1071
Db	1120	CCCATACCCACAGAAACCTTCCAGGTAGCTGATAGTACCCAGAGTGGCTCAGAGGGTG	1171
QY	1078	GAGCCCATGGGC	1090
Db	1180	GAACCTATGGGC	1192
RESULT	8		
AF005370	LOCUS	AF005370	130608 bp DNA VRL 21-AUG-1997
	DEFINITION	Alcelaphine herpesvirus 1 L-DNA, complete sequence.	
	ACCESSION	AF005370	
	NID	g2337967	
	KEYWORDS	wildebeest herpesvirus.	
	SOURCE	Alcelaphine herpesvirus 1	
	ORGANISM	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gamaherpesvirinae.	
REFERENCE		1 (bases 1 to 130608)	
AUTHORS		Ensser,A., Pflanz,R. and Fleckenstein,B.	
TITLE		Primary structure of the alcelaphine herpesvirus 1 genome	
JOURNAL		J. Virol. 71 (9), 6517-6525 (1997)	
MEDLINE		97404659	
REFERENCE		2 (bases 1 to 130608)	
AUTHORS		Ensser,A., Pflanz,R. and Fleckenstein,B.	
TITLE		Direct Submission	
JOURNAL		Submitted (27-MAY-1997) Institut fuer Klinische und Molekulare Virologie, Friedrich-Alexander-Universitaet Erlangen-Nuernberg Schlossgarten 4, Erlangen 91054, Germany	
FEATURES		Location/Qualifiers	
source		1..130608	
		/organism="Alcelaphine herpesvirus 1"	
		/strain="C500"	
		/specific_host="Connochaetes taurinus taurinus"	
		/db_xref="taxon:35252"	
		/lab_host="MDBK bovine epithelial kidney cells; ATCC CCL-72"	
CDS		918..1211	
		/note="putative ORF"	
		/codon_start=1	
		/evidence=not_experimental	
		/product="AI"	

polyA_signal
CDS
/db_xref="PID:g2337968"
/translation="MRLHFLFYRSIEFMYKKSRHKIFCIITLSLYFLTAKHRSARNP
ITDPSFRIWGPCFKWSRHQLLVNQMLSPCRNKCSLTADQCITANEDTTK"
1246..1254
/evidence=not_experimental
complement(1300..1305)
complement(join(1796..2230,2313..2477))
/note="spliced cDNA detected by 5'RACE-PCR"
/codon_start=1
/evidence=experimental
/product="A2"
/db_xref="PID:g2337969"
/translation="MSONSNSNSPRKRRYVKMCDLTEQKERRRSINRASKNFLLK
RRRIFFEQEKLINLYKENSRLRCQVEKRDRIIRLEWLNHKTITLQNYTGPPE
PRVKNSELMQCAELNLDQQTITNNLNPETVSGNNTNGFAAATATLHTNCYEK
TLANNNTNFEAKLNCVLPSTFSAIDLLSISDMNNLYN"
3492..5453
/note="A3: AHV-sema, similar to Vaccinia A39"
/codon_start=1
/evidence=not_experimental
/product="semaphorin homolog"
/db_xref="PID:g2337970"
/translation="MAYLNATVSPVISLSSKVKLFHFCGGGCGCLGLITEFVFIH
PAAAGTLCVSTRLLMILSAITAAKSRFIDKPRLVNTLDGQHRFGPQPHVLFH
SLNSDYVVGNNITYLFDFAHSSNASTALINTSTHTRLSTCFENFTLLHNDP
GLLAGCNSQKPSCLNLTQLPGKLGAPSPSSGNLYLFDQNDTSTINLYAS
LSGSHKRRAGQVELTSDTAMHRPQVQATAVHKNESYDDKIYFFEQENSHSDFKQ
FPHVPRVGQSDQGESLSYKWTFLKARLACVDYDTGRIYNELQDIFWQAP
ENWEEETLYGLFSPNFAVCVFTQKIDHVKTSKLNHKLHLPKPPQCKMKNH
QWPTFQVADPEVADPVYQKNAMFPIQSKYIYTKLLVYRVYGVGWAATIFY
LTIKGIHLYVRYEDSNSTALNILEINPQKPAQIONILLDNLTNKLXYNSEWYS
EVLPLDSVGNDCFSFMSRDPCTWYNITCFQKQRYVETGCPNARTISEMCCDHY
APTIVKHQVSLPLNSLSLCPAVSNHADYFTWKDGFTEKRCHKVTKHNDICILLIANS
TTATNGTHVCNMKEDSVTKLLEVNVTLN"
5590..5597
5732..6097
/note="ORF containing putative signal peptide; detected by
5'RACE-PCR"
/codon_start=1
/evidence=experimental
/product="A4"
/db_xref="PID:g2337971"
/translation="MVAQLYHHLTILIASIYIIFVNAAPTLYEDDEENSTEDYDF
LDSFQFPRNNSCNTREATTLEALKRALQIAGINETDTELTPTPTMT
PLETIPDTPSPVLPSPAP"
6123..6128
/evidence=not_experimental
6333..10442
/note="ORF03; similar to H. saimiri and EH2 ORF3, similar
to ORF75"
/codon_start=1
/evidence=not_experimental
/product="tegument protein"
/db_xref="PID:g2337972"
/translation="MTRRELQLLAYTEPSALETAATAALRSIPGLONVLIQTQDAVL
VTFTSRRLREHQLKETLFLKAALRTDEQQLFIPUSALSSFTFVIGPDINRL
PTSONELTALQTDAREKFNVERIOHCRLLFFSGPGEQLNHTHYALLREILCGD
LTLQPNELNLSDFYATIPYPRFKDKRVHYVHGIDINLTGAASNYNNAASPT
AVTDSEIFRIPYTHVRFIDWEIRVRVPDVGHSRGVFRQHGGAAYMCSTLNST
GATCLOKMGTSFISOPTLGOLGLITEPMCGWQGOELSAIOMLTKYANTVEAFAP
TLNAGVPLVQGVFTLPTLEKRAMAFTSGSLTGPVYLKPPATETLDAEMERN
QLLLEVGYPDYGGKLNPNVIMNSSEGRHAHLHQALQOLMIIITPFGKVIHCCDM
QYNKVTMFIKAKAGDLGSLMSARSLPAHTLRLKKFNRLNRYHRIIKNSWLKYD
SAAALVYADNEYQDQSKMDIALSGWCPFHILGNLTPNSTNTVIGDKNOYGGIVDI
QYKMHKPGQSEGEDTLPAPDMNIOLEKNDVTEBLLQLVLRHPTVGCKAHIVHV
DRCNGHIAQOQGVGPDIPLCDFSVTHNLVDGDIREGMESVPRVNAADRVARLLI
ETQYSTGLDITDATALNGLTTHISEHQVTVESKYGNCIGIGKTIPTQRPDL
GFIYNAENGLRSVVATANAPCVVPGPSLPYFKPGSAILRNLNHTHEFLSGDI
CCMASGIGASGTFPTPSQLNQLQFMLIVKAENALSGHDVSGGLICACENMFAG
GLSARLLIHDEDEPVPPLFSETPGFVLEVNADVAALIAANLYNVECTOIGGVES
DTFTVFNQNTOLLSVPVSRLLKHNLTLSKSDVLLIYKVEDQVLPETSYGNYEVHLYVD
PYSIIISOSTRPNVLLHLLPGCGYPDALLAALTNSGFSPTVTVPYCGKYRHNRRDAA

polyA_signal
CDS
PGSPTADDTAGIVLYGSSNIDSGDSTIROMLVNROVINDRVRLNKAAGSFTLAI
GOLACRILFATKAGDAGSQQPFLLPNASRYESRWLNFKIPEDTKAVAFDRDLRC
VLPCWVGTHLGSFHNNTFFDLETRQQAATFNGPLVQSOPAREYPLNPTAEHPY
AGLCSDEGRHALLDFPCLAFNTQWQHNTQPGQELPSPWKLMPYRLYNWSKFHQ
HYRSLRLNLRHTFNFETQQLDFPHADRRGAVPHDNPQYIPMDAMDPSQ"
11134..11139
/evidence=not_experimental
11260..114643
/note="ORF06"
/codon_start=1
/evidence=not_experimental
/product="major ss DNA binding protein"
/db_xref="PID:g2337973"
/translation="MALKHOLNAVEDNLGSKAPIGCPGFIYYPETHFNFKEISLLG
DKYAEAGFSLPLGHVYEEAFVNPVAVYKKIDMTVTSVKLSFTYNRAIIFHNVEK
FEFISGPGGLGSKACADLFYRLKDRFPVPLQGSRRDPYVITGITVFNELDFLGNF
FLERLHLGLIYIKSHMSHVNKTEYRIPLYDEDLFTKFKSLRRLYLPAVSELYY
TLVTSLAQSLRVHNAASLVEAIOFVHDYKMAKLVSPKEYPLATVYAGCDTTLMDVD
AVAAELGSLYSLSFFEPAQETKVQDYYSHDIFASCTSDRLEALSKNALQAIHIIH
AQLFSTNIIYVNRVRAQAFIPNSKEPVPVNSIYIIOHGLANICEETLFDGSPAFYTG
APASSLDGSSFTLOHLAYAAAFSPNLLARCYIYQCHQKQKSLNPAINTTEYVGSAA
NSPVCSLSCQCPCEINTLFYRLKDRFPVPLQGSRRDPYVITGITVFNELDFLGNF
ASFRKDEDQNETEETPRYTWOLNOTLTLEKLAAGLVDPSPVADGAGGSGNNLEKF
VRFSDIDSVDAAEAAKFINMKNVNFESIKGVSHVIOYNCNTYQWAPCSIMLNL
YRSILTIIODIALPISVYSESNPAQGYKPNWELKLHYOTLWTFKSFIDKGVITG
TEMKVHAEQFSDFDVDAATNNMISPVKVOVRLARAOVALKNIKVKRILFSGTSM
SEHYQNAFLTANRRNDYILAGYVFKFLSHRQPLNKLISCLYNSNCKKQIPIC
VPGVSAALNKFYSYNNNSQFEENMLDVPDYSYTYAKRLNNAILRACQTOFY
AVTIIHSIFPVVKNKYTGACNAOLFOSANIGYFMGVDKNLLGDSLFVKQONSVMR
RSPIVTPVVKNYTGACNAOLFOSANIGYFMGVDKNLLGDSLFVKQONSVMR
KKYLEMTPVGNLLKPSYTHQGTAFETVKRTIOSLEDOADEDLVNRVCELVKSL
GAGCADUTLDDIQIYLGSGMFSENILEKLDQLRELVDGWPMTHEAESVLKSGTCETDE
14725..16767
/note="ORF07"
/codon_start=1
/evidence=not_experimental
/product="transport protein"
/db_xref="PID:g2337974"
/translation="MGQYLAALYSQIYGLDLYSVLEFCFKPTSLCTKIADAKNVHK
IDHEVVSALQONSLEACALSLHLENLKTLPFIYHALLDNPTYFSLKHSVGL
CDLKKLVNOFYNECGEVLNLTINDIERFLNLCVFCYCLSSSALLAKAELFTLG
QLRGISPVPTDIYIYSSSCLECVLETSVYVNOGETINELLNHNCHLHVVPPEPI
KLGFSENLQNLGKHIAITDIQSVGKEHVAIOESLALYKAHTIFNNTPRKVLESLN
LLYNSGQNPQSDSGKSELKINSRENELQYRPKNNGEPHFFDLSPOGTEL
LFCGGIFSTHDTITALQDCSNTFMKQTLGTGAKRQNELFMRLILYGEVPTKP
KOTESALKTCDQSDAKNOVLOEALKEAYLNKLSKEGFRKLQACLSLTHEEMLNSP
SLAKWGSVYKQSNLTLLNHLFRQSWVTOASLPSPVNGSEQFENSKFIKSLLYVKS
SREYLSLRLRHFFALITGLPTOEGLPFPPNPVOLAHCLAAHFMHOKMLLNMIRP
TMPEODMICSNFNEFYTIHETDLNGVOYECWKYLRDLVLSVALYNITWKNLCIYRTP
HSCPTACSGIKEGLYVYESHAPLILVYSNKKWIFKDLVALLYAHMQLANNAHR"
15048..15053
/evidence=not_experimental
16751..19315
/note="ORF08; similar to EBV BALF4, CMV UL55, HSV UL27"
/codon_start=1
/evidence=not_experimental
/product="glycoprotein B"
/db_xref="PID:g2337975"
/translation="MAHTGSTVCAFLIFAVLKNVFCOTPTSSSEVEDVIEPANTVSDN
IYQQRNNTANGIHSDFAPFPRVCSASNIDIFRQTSCHSCPNTKDKNEGLLILF
KENIVPVYFRKIRKIVTSTIYNGIYADAVTQHVFSKSVPIYETRRMDTYQCYN
SLDVTYVGNLIVTDNDGSNMTVDLQPDVGLNSVRSYHSQPEIHABPGDRLTGVYRR
TVNVEYTEDARAVPFRFITNIGDTIEMSPWSKAMNETEFGSPDRLTLTAKDY
RVVDYFRGQPOGHTRIFVDKEEYTLWSAQOFNISYCHWAHWKSFDAIKTEHGKS
LHFVANDITASFTPTQTOREVLGKLVCLNNTIESELKSLAKVNDHSHSNGTAQVYL
TNGLDILWQPLQOQKLLDRAKLDVAKQOONTITTTTTTRRRQRSSSGIDQVTL
AESTILTOFAYDITLRAQINNVLELSRAWCEQHRASLMNLSKINPTSVMSI
YEGYPSAKRIGDVISCHVQVQDQSVLSHRSMVRQGRDKTHECYSPPTFKFINDS
HLVKGOLGVNELLTATTAVEICHENTHEYFOGNNMYFYKNYRHVTPMPGVGDATLD
TMVNLINVENIDFQVIELYSREKRMSTAFDIETFREYNTYQVITGLRDLDTDL
ATNRNQVDAFASLMDGLVNLAVSAFATLFFSIVSGIINFKNPFGMLLFG
LIAAVTITVILLNRKAKRFQNPVQMIYPIKTIITSQREELQVDPISKHEDRLMLAM


```
/db_xref="PID:g2337984"
/translation="WASNSHNNYTPRRQNVDPKAEEDSLYENIDFLTPAFVSSDT
SDEPLPLGREGSGSGSTRPKTGRGLVRROADNPSSDEGDFWEPLEPFSQPG
ARFSEPLERSSNTASGSSCTTILGSSGENPTSHANLSEEMEEIEPPLASNP
SDERPROLGSAGVAGDAKRSKPKFKPAKFKGLSSLLNPFATASDGRARRQPT
COGPTPIFDIGDEPHYKAKCTFEGCMAKGTLLNFARQTLSDDEALRIPEPMRE
WTEVYTKNQISQIKCKCKPGKSTAEELVSCOLKFKATPLKTSQLORSVKDSE
QVGPLDKWVI VDRHQLSALVFPVLVLRMRGLSFDENLLGMFEAHPGEVIALMSV
NVEENFRLKRGRCERHIDRDYKEIKGSFNAAAYCAWFLQYFSIQTWQICGLS
SLDEACATGECVHTTASIRWNMSMLVTLSIIQSNDYTVQNVCSINFFSQTSLUVF
VIDLSAFRHVDVPGANGEFYMQVMKNGDIKTRVMDFTAIKADTAHNTASLD"
39960..39965
/sequence=not_experimental
/feature="ORF22; similar to EBV BXLf2, CMV UL75, HSV UL22"
40059..42260
/codon_start=1
/sequence=not_experimental
/product="glycoprotein H"
/db_xref="PID:g2337985"
/translation="MFLILLCVTGAQALTTPAPRPATTTPRCGVTSAPIVAPASS
ELIVLDGTFHVSVDTEIRQYVQREIEALNASHVFESLETYINRYKDYRTDQ
SIRVNRATKSTCKEVNSTEVSFYKTSISQTINGKDGDLGSHNQQLFFYVMN
VFPENATPVKRVKHYVSSLSADGAYLAGMATTVYSLVVRKISSTVHEATVF
GNKKLLPSMRSQITIDYISLVNSDAEELLLTSQKDYEFYFKNLPQNNQTVFSLTS
HTVGLAQILOTSVYDFARKGRCSRVHSHFTLYLAVLSLYKMGTEFYKNERQI
SIQCLPLKYEVNCFDMVHRCFTSQYTRGEDSDGINRLSAILGSMFEPENQGLSV
TNWFLQTLFYVDCNDQNGLHGLIILMLDIYGRVYVNFILTPEDRETILFYVNALR
GRKHUSTMKNYSLIYCYTSMCATELAWGIEYGEESTSHSAHFSFCFMSLR
DYTLKLEINYSQDKVLTQLSNGVSAMYSLLTAKSTWTIDSLIKPCYNASGVF
MIVPTNYSYISQGVAAAGTYDVAETFLKSMVITVVSNECYNLTAKEILKIPV
VYLMNHTLFEIRGLYANRAMNIILLLFIIAAGVFIYKYIWMYTFK"
complement(42219..42224)
/sequence=not_experimental
complement(42253..43458)
/sequence=not_experimental
/feature="ORF25; similar to EBV BCLF1, CMV UL75, HSV UL22"
43455..44455
/sequence=not_experimental
/feature="ORF25; similar to EBV BCLF1, CMV UL75, HSV UL22"
44565..49767
/codon_start=1
/sequence=not_experimental
/product="major capsid protein"
/db_xref="PID:g2337988"
/translation="MEVCRLENRPLPHAAVEANLLRQVKEAAEGLFSKOLLVGKD
HNSRVREVLGVYANVIEFKFLETGLAASCIINTEFRDLKRMVDGKQFKVSVPTIA
RSDGRPNKQRYIVMKTCKSHHJSAETALAIADLENMHMEPETDMLDLEIGTVKTV
TSALQFGIDALERGLIDVLSVKLHAPPLFILOSLOPTLTERGLKKSISKDLVSME
KDLHTHSFFINKAEALATPROYLGLMSSLIINSVSKETVFKGTSYTYSSGEPVAG
ITEDDAIINKLTLILGEEKYTPQVGAATASYVRGENLVTAVSYGRAMTSFEQKNKL
VDDPVQAGSLDKNADSNESYSLPQTTIPVSVKVGQPVIVISIQMYNEAOAFPP
LNRMYQYTFPLGLFIPRPYITSTSVKLEDDSCLSSEVVVVKNTNTPFLCFNYNAL
TLTCHPRVNSACLOLQRLNADNALEAFRDGLSRPMDMNLFAHRRFYLKRT
VLLPLIAQCNLSTDDLLHPNHLLOLELHPLFDFVVERVAEEAARFATHRPFSGN
IPQLAPNQFODSRGKQFEATSLQHAVDATLEVIRSTAFDPTPYFCIYVEAMHG
QPEFTMNPFIISICINIVWNSGNLAFINFFMVKNICTHMGGLINRDAYALYRKI
LGEVVAIKHAIRCMGABQOLNNEQVYNGLLDQLQLPPFAYRDTFAQLNDRGARFY
IGAKHDNSQAFVHALPDFENANOVAPAMHYTRNTFNWDTPEYITVSGNNDADAADL
EKIYVYMLPVCTHGMCGMGADENIAIVLGYNIPFDTIQFNGDDTVLEHLENGPL
RDIIYASVNPATADMLRMVYSLNCPOMTOVVRVKTBRDHCONLGPQEGALIEHTVM
VNGFVCFGISPRMKQVAKMYPVPFHRFYCDPLVAGSDNHNIQYVNNFSQSLEA
FNAPGGINADYAEWHKAPMAKIIITSCAKSTASLSAFTYMNHKLSPATIVQAKKHIP
GFALTDLRTDEVLAENVMFSARSTSVFGRTVSRREVADAVSEFVYHELATETG
LSYSVVTFAHVASTITDMGYCQNLFIISGNEIYDHADVNNYVARKIGEDAVNRHN
PRALIAGVQLSHRPGLVHGOYATCEVPTPTVADAVYFOKSNSPRGASCVSSDIN
NOERAEQPLYDHSLPDPAYEYRSVNPWASOKSLGDLVLYGSRVQVSSPGIYSPSK
FNFKEMLKNNRSFTLLVNEYAQLAGLAATSCDQLQYVINGIDVFEUQCLFLQEA
FPTUSASHRVLLDEYMSFNTHAPVHMGHYFIEEVAPYKRVFKIGNKVAS"
49784..49789
/sequence=not_experimental
49818..50738
/feature="ORF26; similar to EBV BDLf1, CMV UL85, HSV UL18"
/codon_start=1
/sequence=not_experimental
/product="minor capsid protein"
/db_xref="PID:g2337989"
/translation="MAVDSKIVSVCTSLFTDEVCKLOEKVGCIVPLRDSHSONLOT
VGLPALGIDLGSDYVLMCNLYLSCTTALILEVTTDSLVLKISRNGTYQIKNKHYP
FOWDSNQVCVMPPLFDQESNSVDLSNNFTLLLPVVPVEVAHEALQKLVITNLYLR
VTOAEPNNAARMADVLAQTNVYVYLGNYHSLNLEGESLGAFLDNLATYLCIMVGLL
PRACVRLTLLRHGENELLNVFRMRIPDEFNAAANLNADTVYPMTKIGILLITYLO
TLGSFNLSPLQLVSTVTPENLSATCWVC"
50745..51623
/feature="similar to EBV BDLf2"
/codon_start=1
/sequence=not_experimental
/product="ORF27"
/db_xref="PID:g2337990"
/translation="MVGQRLIKHDDGTVQESITADGWKESYKKKTPVKPDSNPG
VKYEELPVRHVFSSKPCQQLFYLLFWFTYITVILLTVPLPQIHRSYRLMMGP
ASYPINCNRATPDRIYSKCSFSCFTDIFMPEIKVQNETFYPNFTKSDGSPADYSSA
LLWATSFITPNCTNFTVLYSDSNSSQNSDYEVAFVLMLEGLFMRHKRCHPTVY
LGRRCRGAHRWRVNVYDTSYLNHSTCSDWASLSNASTYTPQLGPNCSMANUTEELK
SGEVLHLQOQVEIKPSQ"
50765..50770
/sequence=not_experimental
complement(51644..51649)
join(51670..51675,51674..51679)
/sequence=not_experimental
complement(join(51682..52824,56053..56970))
/function="terminase"
/feature="ORF29; similar to EBV BDRF1/BGRF1; CMV UL89; HSV
UL15; splicing predicted from conserved splice pattern in
most herpesviruses"
/codon_start=1
/sequence=not_experimental
/product="DNA packaging protein"
/db_xref="PID:g2337991"
/translation="MFYVKVMPALQKACEELQNSAKSGKWPVETPLVAVETRSE
RWPHYGLLPGVAAYSTLDEYCHLYNPYIDALTCDLQTHRRVATQVLSDDLCO
QKKLFSCPTSVKAKLEFEAAVTHQALNSOVFLKFTLVNLALNFKNSYSDRS
SHIELFKQLIMHTFFFLYSIKAPELCEFCNIEFKLYFNIDTMDAOTLIDFKQKASV
LIPIRHGKWTWVAIISILLASVDLRIGYVAHQKHVAVFETINTLHTFFPGKYK
DVKRNGTIIIFGLPNKKPSTLLCATCFNKNKSIRGOTFOLLFVDEANFIIKKDALPILG
```

```
polyA_signal
CDS
/sequence=not_experimental
/feature="ORF26; similar to EBV BDLf1, CMV UL85, HSV UL18"
/codon_start=1
/sequence=not_experimental
/product="minor capsid protein"
/db_xref="PID:g2337989"
/translation="MAVDSKIVSVCTSLFTDEVCKLOEKVGCIVPLRDSHSONLOT
VGLPALGIDLGSDYVLMCNLYLSCTTALILEVTTDSLVLKISRNGTYQIKNKHYP
FOWDSNQVCVMPPLFDQESNSVDLSNNFTLLLPVVPVEVAHEALQKLVITNLYLR
VTOAEPNNAARMADVLAQTNVYVYLGNYHSLNLEGESLGAFLDNLATYLCIMVGLL
PRACVRLTLLRHGENELLNVFRMRIPDEFNAAANLNADTVYPMTKIGILLITYLO
TLGSFNLSPLQLVSTVTPENLSATCWVC"
50745..51623
/feature="similar to EBV BDLf2"
/codon_start=1
/sequence=not_experimental
/product="ORF27"
/db_xref="PID:g2337990"
/translation="MVGQRLIKHDDGTVQESITADGWKESYKKKTPVKPDSNPG
VKYEELPVRHVFSSKPCQQLFYLLFWFTYITVILLTVPLPQIHRSYRLMMGP
ASYPINCNRATPDRIYSKCSFSCFTDIFMPEIKVQNETFYPNFTKSDGSPADYSSA
LLWATSFITPNCTNFTVLYSDSNSSQNSDYEVAFVLMLEGLFMRHKRCHPTVY
LGRRCRGAHRWRVNVYDTSYLNHSTCSDWASLSNASTYTPQLGPNCSMANUTEELK
SGEVLHLQOQVEIKPSQ"
50765..50770
/sequence=not_experimental
complement(51644..51649)
join(51670..51675,51674..51679)
/sequence=not_experimental
complement(join(51682..52824,56053..56970))
/function="terminase"
/feature="ORF29; similar to EBV BDRF1/BGRF1; CMV UL89; HSV
UL15; splicing predicted from conserved splice pattern in
most herpesviruses"
/codon_start=1
/sequence=not_experimental
/product="DNA packaging protein"
/db_xref="PID:g2337991"
/translation="MFYVKVMPALQKACEELQNSAKSGKWPVETPLVAVETRSE
RWPHYGLLPGVAAYSTLDEYCHLYNPYIDALTCDLQTHRRVATQVLSDDLCO
QKKLFSCPTSVKAKLEFEAAVTHQALNSOVFLKFTLVNLALNFKNSYSDRS
SHIELFKQLIMHTFFFLYSIKAPELCEFCNIEFKLYFNIDTMDAOTLIDFKQKASV
LIPIRHGKWTWVAIISILLASVDLRIGYVAHQKHVAVFETINTLHTFFPGKYK
DVKRNGTIIIFGLPNKKPSTLLCATCFNKNKSIRGOTFOLLFVDEANFIIKKDALPILG
```



```
/evidence-not_experimental
/product="helicase-primase complex"
/db_xref="PID:g2338002"
/translation="MGQPGQSSHEQSNMLAAVCSISFVNLWHSKAKTAYVWCVRVVK
QEPVNTGSQLSGSPRNSNLYLVSELLSERKMLPDEQLNFTSTIGDDILNFMWEM
ALRAKHPVLSLRHGCNMBAVFSLOLNGIMIKGATLQIVARPLQRAQFAYMWHAK
ASADEGHLVLSLAKTIKTKGTYSNNTGEADPPSTKVARQDQVKLIDIFEPK
DYNLGGSDVTWKLDIVDPDCLVNDCEYRNGLNLFVSALYKKIYGGFKPLPF
IAPPSDLTPSPFSFSPNISIEYSPNFKDFSKTMSQSSLAAILTDCYKVNWELL
ECITG1SVPIPEALRHGIGISWPLGATELNKLMGDLPIYVHIKHPHYLIINFKFN
TRCCSLSGDEPKLARYSTGHLQNLNWSYSLVFSVMNFCNLHNLQWIAFNHSM
FYLAHAQDOVSTPPVAAPVFKTNTCI"
63954..63959
/evidence-not_experimental
/product="helicase-primase complex"
/db_xref="PID:g2338003"
/translation="MNFLOKNCRAFTNMTSGHCQQTATLFRSTLLSGDNTKL
FTPIQSWTCVNIYFIYLLCSQVTPETMFKDIVLSFQANGASSACWLPNFTETRA
PKQVRLPSKMKLITADGSPQWQVYWGIPSLAYASYVGHIAETADWGKVTTETNT
QELLNHFIDLLRIFF"
64502..64507
/evidence-not_experimental
/complement(64502..64507)
/complement(64520..65293)
/notes="similar to EBV BBRF2, CMV UL103, HSV UL7"
/codon_start=1
/evidence-not_experimental
/product="ORF42"
/db_xref="PID:g2338004"
/translation="MOLRAKPASTCSNYTLLPLRVLEVSKNKICVACNSPDPVNSN
GNLNVKDLAARLHSHFAGVILCPVIVASDKYSTFDMYHVIOERTVLYRPQNT
VITEMCCIIISALENCVPSASLIQYLDLANOLFNRHPSRDSIFLLGGVKTLSILMH
WGFQHPDVSQPLRGQSYELYKDLSEFDAECKDMLSMFCKSLGDNWNEEQLEA
FTFNLFYPTILTAKFKSKIIISSEIKAEACLLKDCVLSLI"
64623..64628
/evidence-not_experimental
/complement(65277..66950)
/notes="ORF43: virion protein; similar to EBV BBRF1, CMV
UL104, HSV UL6"
/codon_start=1
/evidence-not_experimental
/product="minor capsid protein"
/db_xref="PID:g2338005"
/translation="MCSIAVHPTIQSVNLFELIQKYAVYVQTLXSSVRNSGTFFRQ
VFTNIYDALSSCTYEDVLQDWTKYTSIRQWKPEKEDNAFKWSTFQSWFTMKMT
IDGIIITKTIYIITKCIASERYIDWATCGVVPVYNORPNSKVQMEKKHLEKYS
SVASSKSLTIFHELVESASVLERITSPISYLEVSTYDQYKCEYKASYONKEIG
VEVILPTIGSGKVNPSVQRLAENVMACHRTMEHAKICQLNTPGLKACVSSAT
KVDILHLEDCKNDPKELMQLLIKLAENKTVNGVTDVDEFTIDYSNKLVDRSK
LFGDVNAEHPSDMLKQVSNVNFCKLTQIINOQFETISKELEERAFPLAKINOIETOL
SKOCEPQGTGGRPYNIILSTLDALDLSQSLHLTNSQVTKGQISVNSFTSOYVPP
FREQLNDHELHEIMQSNLSPIIDNOGORLFVXTQDTIFPLGKTHNLGFDV
MELLICAYCTLSYDIAEYLASRLAIIYDVGKICQPNLSDGSQS"
66940..69291
/notes="ORF44: similar to EBV BBLF4, CMV UL105, HSV UL5"
/codon_start=1
/evidence-not_experimental
/product="helicase"
/db_xref="PID:g2338006"
/translation="MEHIPKSFILNMTSEAKVRMIVQKIKALSEETTSDDPSIDWFDA
EFPDIDLGRPLPFTCTGTAGAGKSTGIAALHONCLITGATVVAQNLSKTLHS
ECPTIYATGFKSRHINLLNRKPOIKLGDSDITAQVYVELSKYWPVKIDLIIGECCK
KKEGLYASSEEARVVEGCLSGPLMTNIIIDEAGTSSSHVLMVAVFLWFFNGWL
NTRYUNGKIPCVYGVSTQTDAYHSQDQHQHVRCEONVLSLIIGNTKTSDIV
KISENWLINNRKCDTPEFGLMKLVLEYGLEISDDMEYINPRSKIMDPMQVY
GWTPLSHSEYKQYLGSLHSALSNSQSTTCTLTCTPCVCEVFDHDAVEYKEAVNP
STPLLEWLRNRYLSNQFIDQDMTIVSTDGEASTKVTATKRVKGSVYNGVTK
KCFMGMGTGYAQFKVLDSSEFLDTHGHEQPEHAYAFILCOLLYNSWYFYQAVQSD
CASYISALANNIPIPEVLCAPAHSEACSEPLNEGEDLFYHKTLPPTAKRSTLANIIA
```

```
polyA_signal
CDS
/translation="MAMFFKKNDRMPFIEGAPRRKRTNFTTFPSLOKHQNAKLNKG
DNDEVPSPSPSGESIDLDISLTPPDGDDYSDVEDGGAEGSDPP1KPVRAARLG
KRCAGLKRFTSTSTASTSSGNDGDCNYAPPAKKPAP1HGNPLGGRRRPELDLSP
K1ENRSDSSSRSTTSSSDSWYVNNKNDP SRGQGNENPSASDSLGAQAAC1HQDHAM
PTTPPVKGEDYPWPWN"
69306..69311
/evidence-not_experimental
/complement(69348..70055)
/notes="similar to EBV BBRF4"
/codon_start=1
/evidence-not_experimental
/product="ORF45"
/db_xref="PID:g2338007"
/translation="MAMFFKKNDRMPFIEGAPRRKRTNFTTFPSLOKHQNAKLNKG
DNDEVPSPSPSGESIDLDISLTPPDGDDYSDVEDGGAEGSDPP1KPVRAARLG
KRCAGLKRFTSTSTASTSSGNDGDCNYAPPAKKPAP1HGNPLGGRRRPELDLSP
K1ENRSDSSSRSTTSSSDSWYVNNKNDP SRGQGNENPSASDSLGAQAAC1HQDHAM
PTTPPVKGEDYPWPWN"
70846
/notes="ORF46: similar to EBV BBRF3, CMV UL114, HSV UL2"
/codon_start=1
/evidence-not_experimental
/product="uracil-DNA glycosylase"
/db_xref="PID:g2338008"
/translation="MEELWRKHWPALPDHQHHPSDYDILLAEPTWQFLDNTFFFRQKL
NFTSKINENMRKSCIIYPPQAEIMSWSLCAPEDVKVY1IGQDPHGGGAOASVSN
KGFVPSPSQNIKFVRSYCDPFSGHGCLLEWRQGVLLNTLT1TVBAKPGSHSD
LGWTFWTNLI1ISTLNKLNKNCVFMWLGSKAIKSLMINKORHLVLKAQHPSPSLAANKN
YSSQCKFLGCGHFTANKYLTOHKNKVPIDWLE"
70686
/evidence-not_experimental
/complement(70681..70686)
/evidence-not_experimental
/complement(70779..71285)
/notes="similar to CMV UL115 gL, HSV UL1 and EBV BBRF2;
weak positional homologue"
/codon_start=1
/evidence-not_experimental
/product="ORF47"
/db_xref="PID:g2338009"
/translation="MMKWVTLTLFVLCGDNPNVAAAHNFCVCHQKNETAHTPPKK
AMSWNLAILRSPSQCNHTNVALYNTTKYQVSCVNGFLMSFCALFDRLTLNIV
RYADOKFYDELLGYKRGFAAQFSKATDSSGFKNLELDLIITRHGRMASKTHVAGLT
RSSASSQL"
72661
/notes="similar to EBV BBRF2"
/codon_start=1
/evidence-not_experimental
/product="ORF48"
/db_xref="PID:g2338010"
/translation="MELPVPVALLTSNQTKWRLLIASFSSSHKNTOACUNFLRGTFCK
TDQCYCAGLLVLVLIQTEDNMIKDRVIQMAILVRSIAEYCFDEIFQVYPENIESM
FTECRRALLLECECGMECLETYKGLQAKQISYRMPRLNPHIESSYFETLSSVNS
AVLCSNRPVSKDLRDIVMGSHGEGFSTEAKKEASLLAICITFCWL1FYKQOVTGA
LEDIDPMQAVTVK1PLESKRDLCLRDVLLKKIKERGAEMGNKNVVKPIAAGT
SLYKOLSAKROLIGDPCFPDIDYENLKNMIQDSAGLTQTPTPCIEGSPDITVQEC
QSVSPDPKGBGDCPHPIINTSPCLKQVSEHLLADALLAESSCPQTQTSLSGVNS
DYLSDPEVTSYENYTLFEFERKINLL"
71407..71412
/evidence-not_experimental
/join(72825..72901,73121..74903)
/notes="ORF50: similar to EBV BBRF1; splicing predicted by
splice site analysis"
/codon_start=1
/evidence-not_experimental
/product="R-transactivator"
/db_xref="PID:g2338011"
/translation="MSANPNCSADPPPKVKTFTFSSIRLSRIPICIDDFIDTADL
GDTI1GAALKSQFNACTQEQSEFTREYDCKNLKIQENKERNEMFGEVADMLNL
FALFYSYQVRVTHFGKQLLCASSQITREYDCKNLKIQENKERNEMFGEVADMLNL
AKEMVLL1SEARGKALNOGRMSGGRONMNAKKVLTVYSSLRDDGETSPVAYMA
YTFVPDLEQVFPQFLKEOETRKKGKAT1TOSLLFAPKRCPCNKTVFSQYQARQVVL
PVLLEASEPQSTLAYGCPDISLLRDSSTOENDESGCCSCSKTSLSGVQPOQSEY
DPSPIPPDSDTESCDQSQRVPESTSDTHAEDDDVPEAPQAASQTPITTTQETQSCCS
```


VYNPATQTGFGYNOQSDNFASRIDASATSTSYGPGPVTNHGSESTVSFHTVPTTIA
TCQQQLHQMVGQQGQTTIYGSIVGGISDANGQSVGASISITISPAISRSSTALQWT
TILFPAPSSASSQSDYFSFGFPYSGSTVPOAFELAPSTPSLDELDRDSGLVS
QQQAPFPQDQGGPQYVPVAEQQSQSTIDPLSDEMRRRIEFEEFVSVPNVRP"
75159..75164
/evidence=not_experimental
75226..75858
/note="position similar to EBV BZLF1"
/codon_start=1
/evidence=not_experimental
/product="A6"
/db_xref="PID:g2338012"
/translation="MHKISATADMLPISAFSPLKGLGVVATPQGPVLWDATTTVPVF
QAFPCPALPQVFNQPADAMDKIMRPADETRGVIMRPALPEVEFETSLRPAQLE
LPEPSQSQSDVDDTOSKEDTETLECAQALTDLKWGTVDPPRSTSPVASTSSGV
SSDLQAKAEQAIRRYRLIEAKDLGLVVRGPKKDRDPDFVIKKVPLHFS"
75865..75870
/evidence=not_experimental
76569..77300
/function="unknown function"
/note="positional similar to BZLF2; structural
similarities to ORF51; contains predicted signal peptide
cleavage site"
/codon_start=1
/evidence=not_experimental
/product="A7"
/db_xref="PID:g2338013"
/translation="MLAEMLPVNMMLPRKALLVDIFFILAAATNLMTAAAFALGCLAF
YKQLVITIGNITFPHQSGDEVIRAMYIPVNDSDVDENPGFRLSNLTLPSLDGPDY
SNSQCEICPGRGVGKACYVPPTKYSFQNCFFACKNISKCFYLTPQNLIDPFDDHT
LRQDQWIGTFKKKNAALSTIDNNDFITAMDELSYCAILTRRSRIVYDFCTTISK
LCLCGQEDFTAPFYEALTAPVHG"
77293..79344
/note="A8; position similar to EBV gp220/350"
/codon_start=1
/evidence=not_experimental
/product="putative glycoprotein"
/db_xref="PID:g2338014"
/translation="MDNVTLALNCNHTTHGYAHYVYSTLSYLVFIQNGAPVNCBNCI
YDITFNSMLNISNNYFIANATILGAPEYPOQVTVLSVIGESPVAVWLAVHTPDKN
YEDTARAVNRVLDPENISWCAYNVQSVDPDTQDISNCTEYSEELPVNNQKSSI
HLNTVLASTVSVLSNASNITSQSCGNSLENLTSNVHIPCPLTLTVSKASTDN
MOTGAANASIDIGMLQNLQDASLTIVSSPGLNLTQKNTIYKNSKTDHRVRDENLY
VLAQIPALEGHKDSITRSPILNFSNSEFQVLYDRNSSCVWNPANGDTFIFY
SMSCPHITATISPRGEIKVNTGNEFSRANLSLAFLSKGRKEYAGVLLQAHEPSTRPP
PCTVAPGILSTANFETSTNKSPSTPTPAKLSPPGLTNTLLLTAGEHNSGIGSTL
EPLTVSVOLQTPSSPTRDSTLVILKLTTPQDHTKTVSPSLVTPGRTSLPIVSMTH
FSREGSKPKPTAAKTSEASLPLLATTPPTNTKESQSTFASSTVSDTFTGDD
VNTVTMSPSITQTLPIPTSGRQIYVGCCTLNRSGNLFFFELFVAAHRESGHNTS
MPNHNHNSVKYPEDHPHHPGDHDDHHERFOIWLPLPIAGTIIFALVALVIVNIALRMT
E"
complement(79385..79390)
/evidence=not_experimental
79388..79393
/evidence=not_experimental
complement(79398..79775)
/note="position similar to EBV BLRF2"
/codon_start=1
/evidence=not_experimental
/product="ORF52"
/db_xref="PID:g2338015"
/translation="MDPKNLVTEQLAAELTKIQMENSHLKRLKRSVGGPPKPPKPPR
ELTEPERQLVLARHNRHRSSELLRQLDKLTATLVTEEDIVELKNADFLRHP
DPSNPCKLSKRRKTRARGQQ"
complement(79845..79850)
/note="speculative"
/evidence=not_experimental
complement(79854..80165)
/note="similar to EBV BLRF1, CMV UL73, HSV UL49a; contains
putative signal peptide cleavage site"
/codon_start=1
/evidence=not_experimental
/product="ORF53"

/db_xref="PID:g2338016"
/translation="MASSVSTCVLWVLLINWYSITSCDSSSTPRAVTHPVNLNATSNFN
PTAGFYFSCNADTYLLRLNSFSFSSIALMNVVVLVSTVFMVTLCTCFKFTHTLIYQQ
K"
80228..81124
/note="dUTPase; ORF54; similar to EBV BLLF3, CMV UL72, HSV
UL50"
/codon_start=1
/evidence=not_experimental
/product="deoxyuridine triphosphatase"
/db_xref="PID:g2338017"
/translation="MSRTLMSRRQRLEMYKKVPSVFTVTSNOEESTLQLVNKNKPVV
IEPRSSVPLGVYLRCLPGYACMLLANTYRNTFHRGLIDPTMGEGLKLCNKRIFA
IVVYPAGRLKTVIAFTLSPLIPGPSVLSPPTDDAGIDCLDQLVWVPLKRAFTD
QALUTCIOQKNFTPVVLRGSLAAKGLSITPCPKGDLRLSMFNMTSETIILPEGS
RLCQVFMHNDHLETKIPRIILAAFLFQHLRMDMPCQSRVSFIDIQKDCPCTSTLFLQ
DSTGNSISDATRGSKGLSGSI"
complement(81129..81134)
/evidence=not_experimental
81138..81163
/evidence=not_experimental
complement(81162..81824)
/note="similar to EBV BSRL1, CMV UL71, HSV UL51"
/codon_start=1
/evidence=not_experimental
/product="ORF55"
/db_xref="PID:g2338018"
/translation="MASRMWETCCGLWLFKGTKQKKTQLTQLEPSPSKCSERVKQEIN
KGLPGVSGDLIIGDKSTEALNOAHLALQSNITVEYLARENAEIPASCQGVSNQ
IDKLKMQSVITWNAKISATSNVELSDSGQLLDKQACENKTMLEKELATLSVDN
TNNWARELSNIIITQPTHALPEAVFERPEPIYDDPDEESTMLLQPTQOKKATVITQQQ
IL"
81823..84336
/function="DNA-replication protein"
/note="ORF56; part of the helicase-primase complex;
similar to EBV BSLF1, CMV UL70, HSV UL52"
/codon_start=1
/evidence=not_experimental
/product="primase"
/db_xref="PID:g2338019"
/translation="MTRVVKALFAADTCAELLADITGQNNSGSLYCLLHNCRPEDV
SWETRLSLCLPVRPGGAGCAEVEKIKLPTAEAGQFENCGSLSNKOVYESYKR
AAETFPQILEVLGSGDEYSIKNAVTWGRKFVASKYKFKLTTSPIHWFINTGAFE
NHFVLYSCYFFFPVSVTDLCHAMLYSSKKGCPLSITTLREGAVATSPALD
RVENFYLYCDKLARDLECAVQRCNEFRQGLMSQDLVHYIYLSFFQFNQKF
LAYQCNPCNLNTMLREPLMVAIDSDFKHMATYKNTYLSNVLLRGHILHPV
VYGOECLSAHATGGHSIMWGESHOVDMKLTINVEYDILCHEEERGLMDLAITD
RCSIEGPNPIHLVTDCAATSGATPIYRSLSLHSHRYELAVESDDIEYFWKKTIFLPESEC
LGAQDTMLTRAITTEMHCSMSSVAEQIHSRHEYFNPKLVPFNWLDLPLISEGNL
HDSIYSLCLLIRSEVDLILKLLGPVEPDHEVFFKSAICINLCPGEASWRPSTFCIC
TEKGMRVYTRFPFGICLKGSEPLTQTKILNRVILKGCGLSLNLSFQLSNGPFDVG
IYGRGSRILPHTYKVGKQGLERLLKLFVCHPEATDKFPYQLNSLKLNRLLHQAQSQ
DPGRLKTIYVEDINEDFLYKHTQKOLPVKHEAVIPSIERLLDTSLNCFLLSKWPK
CFGTIRSYMSEKIQQPSRVVFPHTNHCIVQVRDGNNGNFKLCRLYNHRGSSKSVRVFL
ILHLKEETKLIVTFMSQCFANKQSNKAMAHFSVFVDLS"
84370..84375
/evidence=not_experimental
join(84462..84513,84605..85863)
/note="ORF57; similar to EBV BMLF1, CMV UL69, HSV UL54 and
H. saimiri IE52; spliced cDNA detected by 5' RACE-PCR"
/codon_start=1
/evidence=experimental
/product="transcriptional control protein"
/db_xref="PID:g2338020"
/translation="MAQQAIVTMSALRRTMEVSDSDVSDISIAEDSNDSPHLEESVD
DCMDCKPNNRPNPISMKPARRVMPKRRERSKPTQHTSPLNRLPNVVLGHQGY
KQRPAPSARSRRPQISARKDSAAKQPSQNPQLTELLKNVDPALASRTIEMKRP
SMLTPSQGFAHMLPMSAEDSKFINVPYNVMEVEHVNVVVRCTEAWLISLSLQD
KSISTKYLAEFNFDLRDPAQRSINKSAWLNIRREAIANAGVNLCAFCADMMMLQLN
LNNQSWKACREDIILTGAFCMLQKVRATKCFLRERHQRALVNALCHITCEFG
GIKQAATKQELFFDFKGLMVLVFLYAFLYSHITPQCNGGYSKCAVQYTPGAV
TGLNSAIEDHYKDCCTSDCTNLITAIYVSPITSNKGLLFFFLPMP"
85957..85962

polyA_signal


```
polyA_signal
/sequence-not_experimental
complement(85977. .85982)
/evidence-not_experimental
complement(86022. .87077)
/note="similar to EBV BMRF2 and HSV UL43"
/codon_start=1
/evidence-not_experimental
/product="ORF58"
/db_xref="PID:g2338021"
/translation="MTKSDVQOLIAAFTSGLLASSIVMSYIEATVFSSSMLTWS
VLYVWSPIVQLAAIFACAVRNFSRLGILLNCGIAFLFSISLWSISLVPGGLF
VINLSIMWIDCFDYYLCPPIYHKIFELGFLASUTIHCFNOFEIYITNNMFAP
FICMLFGYIGSHVHRDMYIFGMRCKPIYITKRAKIIAFSAWQIMDVALFICLW
FLMLAAGCALVMESEVGLVSTYILFVMGNGFCGSLIYRSYVMTAVYSVSLT
AFFLILMGYLFTRKAQLSMLAAVMFVCFHANGCLLYRIKKLHRNITTPRILNVCM
LLNALLETVLLAOKL"
complement(87084. .88319)
/function="DNA replication"
/note="ORF59; subunit of DNA-polymerase; similar to EBV
BMRF1 (EA-D), CMV UL44, HSV UL42"
/codon_start=1
/evidence-not_experimental
/product="processivity factor"
/db_xref="PID:g2338022"
/translation="METCASPLSPALLTSSKIYDHVSKLSAMIQFSDLNTPVM
SIINMGSGAGIVTQLTGAVSDETETLAEITETPLAFNRHSGFTYLHRSREFGKEIEDI
LVRYIKRASTASKPEVETVKITYNNSITETRHSTVDVSHVSPVEKYLQKCFVKAKLI
LSIKTMIQKWLQKQNKSPVABHINEILTVLVTVGDECTIIEKSFVLEPADAF
LTUKPQNGFAGVDCIAVNLECIQAIGICKVPSCVPAFRKYSIGGIVEVSSAHLK
QSKSPSTAVTLDLASESKLPQAEQPTTSPQSDPOSSNSVDKPPRPSPISQPVYR
RRAPPRAIISDSESDSDQDCSTETQPTSETREBPKEPTPRPTAPPNKRKQFEPL
KCPVEKKKNSNSFVSI"
complement(87825. .87830)
/evidence-not_experimental
/note="speculative"
complement(88430. .89347)
/evidence-not_experimental
/note="ORF60; small; similar to EBV BMRF1, HSV UL40"
/codon_start=1
/evidence-not_experimental
/product="ribonucleotide-reductase, small subunit"
/db_xref="PID:g2338023"
/translation="MDRIKLYVCDHPGFELTQETFQNRWRPQAQINLSVDVKCLSL
MSEADVNFYKLTFLCMAETLVNFIDELIVDECHDIKHICYEQAMCEVHKVYIF
NINLNFKNLNAETWEAEAVLKDEALQKLEWESIKMAKTKAEKVLIFLLEGVF
FISFYGILRVKGVMPGVCMDANDYISRODELLHTRAAALLYNTMIPKEERPSSEWY
SLFEAVEIECAFTKQVTFVSMDDIRAFLEATADRLNLTIELPRYKSDPPQSC
PLTYTGCKNVSPFERESTESYSSFIINDL"
complement(89344. .89349)
/evidence-not_experimental
complement(89378. .91720)
/note="ORF61; Rlarge; similar to EBV BRF2, CMV UL45, HSV
UL39"
/codon_start=1
/evidence-not_experimental
/product="ribonucleotide-reductase, large subunit"
/db_xref="PID:g2338024"
/translation="MAAEATPQOQIDQLKYNAGWDMKSNIKAGLLHHEMEKATKT
VTEIERFADVEENLRFQDNAMLELCIYMYEQLPAYKAVKSRGILSAKRFYDTY
VLRTADGSCVESHCFMRIAACFTVOVLNLSALKITLVLGRDKFLDLPSPMTDL
FIYFSPSLHQVCCATPIMRSAGLRDANLASCFLINPDLSTESTTALLQELTMLL
SAKSGVGNVSFVDEKQIQSCVGLINSQVEFFNDQNPVRPVAATMEVHNSIQIF
LAVKLPNGRSCASIIQGLICPLKFFEFKIDEPQWYLFKEPKSGNLNLYGDEFFS
EYERLIGYPCADYIPKSLMFLINTIKTGPSYIYKEACNEHHKMEGCAIASA
NLCAEVQYPCADYSTCNANVCLPMLCVTNSDHSPLCTYCNKYEIOESQVGVGF
SMTLHAAVEAVFLNCAITAGGKCTPGMERGORESMGIVGHGLADVAEMGYSL
DERAARDVELFENMYRAVKTNSNIRLGGREFGWSGSKLRGHGFHHQGWEDVNL
SIPITEMEKLNRRCISSGVNSQFIAMLPVGSLLTGFSESYIYFANISSVKVSKE
EVMKPNMTFNRNWSKEDLDTVRFFSGDVALFPEPLKEKYSLEFSAFYCAEKQARAR
LRAPFQDSQSHSFHLKEENVVSAFLDKJILSYTLGLTKITMYICKVKKQSTMSFSQ
CLRDQNKSEMTGNDQVEPSYIKCTAAGGETSEACLHCQ"
complement(91736. .92740)

/note="ORF62; similar to EBV BRF1, CMV UL46, HSV UL38"
/codon_start=1
/evidence-not_experimental
/product="probable capsid assembly and DNA maturation
protein"
/db_xref="PID:g2338025"
/translation="MKTQDPPOKKNYDSLQSNLFRLIPTTHKISLARPNGFURGLA
DLVKYSVDGAESLTFQPGAWDAPYVOPAFEDFLVHAKTISKHEPVGPLPCFNKST
APSDVLVLTISSHAAVGLPADVDPNVHRAHIWYGDSEVSSLMEDLNLLESNLH
TRHPVGLIVENDSSFLNRTHALTPGAPMRSKQAALKLVPTDFLDVDDARUNVEA
YGAQSGGTSTFCTLVYTCGNDIKPALTFFFSKNKSDFDVLTILIRAYADLIANKLE
VNOQCNINGLKFGVMCTVGYTDSHSHNQSLCIRGSSLLVTSISNFVNYTGRVFA
"
92739. .95597
/note="ORF63; similar to EBV BOLF1, CMV UL47, HSV UL37"
/codon_start=1
/evidence-not_experimental
/product="tegument protein"
/db_xref="PID:g2338026"
/translation="MMENKRSLLKITQIONVTSKLAKREILDLEMKMIDATELCNQ
OLVVKFLNSLPQGGGLIEVTHYVYFLEKNCITLSPQFLERGTEHEVYVSKLHLK
SAVECMQIITSLDFGCGSGLLNRRNLSLSEQLFLTWAQNNRSTELTSLPTPVOT
FLCAEIMSILLYSKLNTIPSVQFNMPNFTQMGVLEQMTISLTYSTVGLTPGYQLP
NATELACALVSRHADLFISPTHVSQPLLTFFPAKRAANLIFQSYLNNPOASITETGPL
VOLREADLOLTDTTFFLYDYIFEALSNQAYSCSAATVNNFTARCVDLSLDTLGSNLL
EASTASRPNNDTAKFGLLRAGLTKKNSDEKTLNLLKNSGTQWOKFPLRLV
TQULTAHYVACLOQSPYTSLARCKITDLKMAAAHOMVEYQNGSGSKSPFEWTI
PGILSFPIQPTIELQVMSDNISSPYQSFVISAHRAWQLKPHITLQRELSPTPLP
SSPNEEDVKKCYREIQGDATYQLVNIVKCDTEFEIKTHAFPIQNIISMDLHIHR
AMQVRLNITFAADAPFLTLRKLILLYLQINDIISONRVDTSFINMLDITKDVITL
AQVEASAFSPNLINLYLTHFSSKLIMOTVNEFTTETTVASVASLARVAGAT
ICHSVFSFNNDTLEPIAGENGDEIFQISVPFAKSTVANLQNCQDVLVILNNESSR
ELASVLDLQVISEDSDSMRHSIKEDHASLNFGKMDFYIKCFALONKLASKIANVC
CYSLTRFAPLFEPELISQTVENLNFSDTVDPSVFLSGINQPCMCVPPNCOGAV
EKLSQVDLIRELAPDTEGASAPANATVIKHNFTGTFDKVSDIDLDQSSSLNTEFTF
95612. .103432
/note="ORF64; similar to EBV BPLF1, CMV UL48, HSV UL36"
/codon_start=1
/evidence-not_experimental
/product="large tegument protein"
/db_xref="PID:g2338027"
/translation="MAFOAQDTDTGLQLLATASHHQGCKYGPAGSOCLSVCVVYLAS
SFWNAPVDSREGDVLVLLHGSLDLRLMTHNETPMNEFAQLSSVPKVLHLMASCAI
ETSPELYGLSDECVCAPFIMSKALENNYEIPQIYILICNGKSGAIIKNNKTF
IFPHCVSSRDTAAVYASASASSVAYVGEASREYTAACHLYFPAESDSEFPNSYLLA
NYGIASALKRSGAYVNLKTLQTHAISSTQPIAPPVPAAPSGMQFFPTQASPPSSQF
VATSTPVTKALPTPSHOPISPGAAKLKALHMTVKRKRIASPLHNSITSDSDST
GKNTPVKITPNEADVGEVETFWLDDDIISTKOMLSETSSSEKNLPESIFSSDMDTD
SLISTHESPATPSLKGQVNYTSTFSLDNEELGGSNLQDQTKYKLDSTTFENLPOL
LTSLEHIERKYPHKNDMPVVDHSVNCYREAVALYITDLSLHSHVIOGLLSKPE
HKSOALNIRYIAIWLNKLISYTNVOVELINELYPIYQALFATKDGKLEAMLEIE
KITKCFOTLHGTSLQDMKNLSKMQIASIKNTPIYDKRANVQVEETENLSKVSILPYFIT
TEBELKKIQAALANLKWVIODHNSVSHEDSEFHRAVNSVKNFLPIPSDIOPLDYTL
KADYLDGTVQQLVASTQSCOTISNHLSSMOSNYSFSADIIDFVSLRCKICTLLNN
IQTSKDLGLYNKQLKARQOLAYLGEISSLSSMOSNYSFSADIIDFVSLRCKICTLLNN
KIFTNQOQMLQNLILDEVESLQAASEQTQGLQKLAWLPLSIAYLENAGTLIG
PEGNAQFDRLKEITDILGSDVMLAYKKNISARTLATDQATISNFTDPAKQRFQOQ
ALTKQKDLFTQVNASLKTDRPPTFTENDMAALESVAMNSDSTLSATAAFYKTKITLL
LKKPCTVPTVIAHLKSLATANINSTKRSYLMYLAOLNKETYSKKGLEOTOAPA
PKGPTADPTVDSKLKLDOSONDRHQKEKPLKPKQOALISLPVKQDTHLKPSPON
LSPVSTNKAPVESPESSPSTSSHSPVSNESONGSFSSLESSEELMDTFVADD
HEUSDTHHMDVKTTTISVAMDDKPTSPPEPSSLSAIVQOEEERGEQAWKKIQIA
FKTLDFTDITSDWHVTGTOHDLKVSFTGFTLESLSLQKLNKTKITGAALS
TKYQAFKWPISIDMTPYKNVLYSTVAVYPSVLDAEQSRAEIDVLLQRSIDALL
KGTAGTYLESDSRMLDIVSSIIEAASDYKLVNVEVEKTHOVHNTVNGTDSLPPKP
DVILPSHLLNPAEQTIASLNKDFRYVTVVEKHLAELNSLQLLKVLISENGFO
AWKEATANKLQLISLQSNAPPEVRYQTVSHDPTLFDKLIHDSVSSKLTISEAS
AALHVESTCAHATQCPYAGAKQAIIALVAKVKPLESLSLVSNQANNTDINI
IKQIISLDPKRITGGTSKVQEWTKVKLELADLEASVYQNLQRLHMLTAR
STNLANLQKTTSLYKWKHEKHTGPTITGKELDYLTFLKFLIEYETNOAC
IFSTEALPASKIDAKDALTPLTQSPVDEFNTSGSPSPPINQOOLEAMCHLNAPT
OPTWLQVFTVDNLSDWYIPIKKNASPLSLQVIFNNFIETVFOAPQPGQDTSQYRGS
```

TVLPNVLOAHFGIATHLNSHNWNLNLENSTQALQSYTTVHSLHGTAQNEMAAIIVA
VHGLQSLSYIDTTPANSDDIPVLSFRATLEIILWIPKIIIFYELKMKSFQSGVSE
LQIMINCLINLFAVQIHIITNSLIMAGVPEPNGLFCPKYKWTLEPOAYLWGDKEAF
LOCSNQSKARVCFALESINPVVLGOLWRSKLPTELPOVHTPYDFLKVLEAAY
KPAYDCQVYSKARVDETPTPVSYGPGNSVLRVESKTKKPGVTTIPISGFEMAAAIL
QKFNPIVYITKTPVSNBETGIFVYVGLLDTGNEEPFSSLLTAPLOPVSVENTILA
SLYTKLEKIDFRSKLWLOHLSPPNNLSHLKHPVILLDSERKLTGAYSLTNNAAPPQ
LNLRFAYDGKLPNPMELLOQASTLSYSHQALEDKTKDWIIDIEDLDDSLCIQNM
SAYPPIIHGSPDSDESDLSPTSPTHLPKAEFTLPDPSKPPQDKKSTPNP
PRPTFTPLPCYLSAEVATLPKIPISIKPSARVHVFPKSNLAADQLLRDFESLS
LQDRPRPKDGPDIINVLPPHGSIYNLOPPNPRMAVIGTLRAADOKSVETPRPAGV
GSRKAVYBQLPHDSEVENGDGPPGCGSEBKYSIRQESVQDPTVSILLRQPPHII
EAVLPKDPQSLTAHQOQAFKQLHVCESKSHRKLULLSAPPTIAPNHLSTHEILADRA
SLVAPRPPIIDENLIDIEPLFVQVIMEVSIIEBAKNVLTITFKKIRQAVTHNAQLLAS
SIQRLAALX1"
90675. .96080
/evidence-not_experimental
complement(103341. .103346)
/evidence-not_experimental
103457. .103462
/evidence-not_experimental
complement(103473. .104231)
/note="ORF65; positional similar to EBV BRF63 and HSV
UL35"
/codon_start=1
/evidence-not_experimental
/product="capsid protein"
/db_xref="PID:g2338028"
/translation="MAHARPKPLRINLRDAEYPNDPRVQOLHLOLNNPNVANNVRA
PLSYLFLTAQEMYEVFVQCARGVNKKRSGGPKPLPANPGAGPNARCGGGGSGAC
GGSGPGGSAQLOPPNNGVPSNOVSSNSGAAGSGSGASSNSGAGSGAGGSSGAC
PSSGAGGAGSGASSNQVAPPPLPQVPMVGVVNGQEMYLPLGPNGLPQOLG
TAGDSNGGQAASPGSQNPPTKPVAPRGKLGKR"
complement(104017. .104022)
/evidence-not_experimental
complement(104194. .104197)
/evidence-not_experimental
complement(104304. .105617)
/note="similar to EBV BRF62, CMV UL39"
/codon_start=1
/evidence-not_experimental
/product="ORF66"
/db_xref="PID:g2338029"
/translation="MPVMGCSVCSAFAELSRWAHGYPGGQTDAVAIEGLDWLNLQIL
DVLEIDHAFNVNVLKGLTICIPWGTGREGFTTSWQGLLERCSCPSVDIHKQLFYIG
IYWCYFLTVYLLLPSPVITKYARSEFRGEOLWHLLCKFEWTEKMEYVFKINFNHP
VINKINEDLESYMLBKLLKROYLTLPOLAVPPLFTRLPDLOFIDEGHVLTHGCEAL
ASALKSCCEDVPCGSPDSVMKNLARCALSHOFSVIPVSDSPNVIQIRKILIS
VLACVVRPILSATVMSLVETKRPTTFVYCGECKHCLNFGKGLKLVNPNHTAFYCR
DOKEKQNCVATTGRIYCSFGSADIHTASLTQMLDGVPIIRAVMANNAAFMLDTAQR
SVDFILPCLGTHAKCEGSLVRLSLILQLLYLTINASELMCARQS"
complement(105130. .105135)
/evidence-not_experimental
complement(105521. .106312)
/note="ORF67; virion tegument protein; similar to EBV
BRF61, CMV UL50, HSV UL34"
/codon_start=1
/evidence-not_experimental
/product="tegument protein"
/db_xref="PID:g2338030"
/translation="MASGKKLIDQLCSVVSSFICLSDIDRCVAGPHIFSRGSSQ
AICTVLLHGEVYNLFVRYMAHILEKYNFPESPFTIICNNGLAVTLKCYVSEPRDL
SSRYQAGTSMALDQNRNSFVLSQDDFKTKTPLVFAKDLIDTNSMVVCRTILTS
RNSLOFLVVKSNRPLENVDLMIKRAVEATGNSLPRATREKPLPLETQLESTLPS
GHLRVQSTSLTRCPSWGNACALLLLSLAVGLMAILAAKLQWP"
complement(106319. .106573)
/note="similar to EBV BRF60.5, CMV UL51, HSV UL33"
/codon_start=1
/evidence-not_experimental
/product="ORF67A"
/db_xref="PID:g2338031"
/translation="MDLSNLLDPPSILPEDIQVIAPTSAKUSLLSHCOHLKLFIQQQ
GSKGCAHSRVLEEKLEAVKEVISKIITETDKILEKSEKL"

106760. .108166
/note="ORF68; similar to EBV BFLF1, CMV UL52, HSV UL32"
/codon_start=1
/evidence-not_experimental
/product="putative major envelope glycoprotein"
/db_xref="PID:g2338032"
/translation="NSSNKTSSFVPMKVATILKHOKTLEAVIQKAFLPQDPAEALNSS
QFCETAAIDTAPKICOCLOLCTHHSFDLSFYGDYAIICYIYALHAPKTNLML
LADCELTLQLYFPDAPSPPNINGLDIYLVHFNRCFRLANTEKPMENSLDMLKTEF
LRATLSGSLGACFTKWPSTUAPVRMADECTCTPTINCGDGGKGLRFPACIGK
DNFLDILIFKWKNTAMTPANSLLADTILSRHOVYFQNLTPVETNASLDRPALDTQGL
PCLLSPALCKKKNHTSSLICELCECLASHSEASVQTFKHLVLSINKNKVKLLDRIL
FQQDADLSFIQDRELLKSVLNCSPQEIHKHLCDFCLCALNSLTSVILFGEVPD
FEETAKATLATGNSLVHRFSQCEILETLILLFKSLQTVKANKTTSVIEKVDASL
KKHFSLLSCYYFTIYT"
108189. .109031
/note="similar to EBV BFLF2, CMV UL53, HSV UL31"
/codon_start=1
/evidence-not_experimental
/product="ORF69"
/db_xref="PID:g2338033"
/translation="MHKIQKMSCTPSVRSRYTLKRRKLSAKSATLKKKKVFLNSEF
FAGVSTNYELGKDFLRMDPTICTSNTVFLPVKESDVARGCLTSLPGHSSVLQFHC
QECKPSSGFTQAQSAESNELLSVNLCLFNNKVKVQHKAFYLSLGHSNWTVKQS
LGQPSLLCYTVTKRFPPIPTFANGPMLTMYIIFTSLTHVSEAVLRILITDNVEN
HNLSDACYKGYILSIEPQALBESNLNVCTVKICDLVAQLDFSDELKQEVNGSTLIA
NFLN"
108241. .108247
/evidence-not_experimental
109087. .109092
/evidence-not_experimental
114100. .115820
/note="microsatellite; variable region in several clones,
exact number of repeats unknown"
/rpt_unit=999gacggggccacccagccagccccc
complement(116023. .116028)
/evidence-not_experimental
complement(116376. .120278)
/note="ORF73; similar to H. salmieri and KSHV ORF73"
/codon_start=1
/evidence-not_experimental
/product="putative immediate early protein"
/db_xref="PID:g2338034"
/translation="NVLLRSLGLTRPGEECDGSPSTRTRHGCLPQPNIKSAAGIGGK
FPSPQGRKKRKGPKKSGKKKRRVTGEPGGGEGPGGEGPGGEGPGGEGPGGEGPGG
EGPGGEVPGGEGVPGGEGPGGEGPGGEGPGGEGPGGEGPGGGRKRKRGDGSKKHGK
KKKTVTGGSGEGPGERDDPDGSGQEGPKREGPLPDGPGEGPEGEGPEGL
EGPKGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGK
GDSPDGAQEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEG
EGGEDSPDGAQEGPGGEGPGGEGPGGEGPEEGPEGEGPGEGPEGEGPEGEGPE
GGEGPLGEGPGEGPGEGPGRPSDPAQEGPGEGPGEGPGEGPGEGPGEGPGEGPG
GGEGPGEGPGEGPGEGPGRPSDPAQEGPGEGPGEGPGEGPGEGPGEGPGEGPG
AQEGPGEGPGEGPGEGPGEGPGEGPGEGPGEGPGEGPGEGPGEGPGEGPGEGP
SPDGAQEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGP
GGEGSPDGAQEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGP
GGEGPGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGP
DPGPGDGTGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGGEGPGG
SDSANRALGVCVGSKSDSDDEEEEEEVLIITSSGDDGCGSSDVYCVGEKKGEGKKG
REEDDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEE
GGEG
GGEGEGEDKPPFCRSPGVSGFYDLTWSSDSRSTEGSRGSPGDDLDGRPGSQGPPTL
SQPGPGSGYSNVDREPPVLSPQCGSGNDESDPDSSREPPDLSPONPPGPD
NGNESDPSYQPLAGSSSSSDDDPGEGTSQGPKRPKRHHQPTKRAQGKTLGLDPL
YDPROKATFSLHLGCPKDLPLVRLSRMIRLHPGEPHSSIFFTGQGVVYVYTSYF
EAKLKLDFIIRQENRNPQQGRVNVSLARHYPPFPFPPE"
117180. .117520
/note="microsatellite"
/rpt_unit=ttckctcc
117632. .117806
/note="pyrimidine-rich region"
/function="unknown"
117980. .120065
repeat_region
misc_feature
repeat_region

polyA_signal	/note="microsatellite"	Query Match	20.0%;	Score 380;	DB 17;	Length 130608;
	/complement(121230..121235)	Best Local Similarity	52.3%;	Pred. No. 8.5e-68;		
CDS	/evidence-not_experimental	Matches 864;	Conservative 0;	Mismatches 785;	Indels 3;	Gaps 1;
	/complement(121296..125243)					
	/note="similar to ORF3 of EHV2 and AHV-1, and ORF75 of all rhadinoviruses, and EBV BNRF1; also similar to formylglycineamide-synthase"	QY 109	TTTGGCCAGACTGAGCGGCACACAGGTGCTTTTCCACGAGCGACGCGAGCTCCTCTGTGTGG 168			
	/codon_start=1	DB 3762	TTTGGACCCAGGACACACACTGCTGCTTTTTCACAGCCTCAACTCTTCAGACGTATAT 3821			
	/evidence-not_experimental	QY 169	GTGGGAGGACGTGGCGAAGTCTACCTCTTTGACTTCCCGGAGGGCAAGAACGATCTGTG 228			
	/product="ORF75"	DB 3822	GTGGGAGGTAATAATACCATCTATTGTTGATTTTGTCTACAGCTCCCAACGATCCACA 3881			
	/db_xref="PID:g2338035"	QY 229	CGCAGGTCGAATACGGCTCCCAAGGGGTCTGCTCTGGATAAGCGGACCTCCGAGAAC 288			
	/translation="MSLYLLPNRRVFKDSLRIFWVDYDYSIEEQLVMKFMVPGD	DB 3882	GCTTTGATAAACATAACTAGCACACATAATAACCCACGGTTATCTAGTACCTTCGGAAGAAC 3941			
	LSLIGTIFSELIIIVTETSHTSADDESSYERAAARELLSPMLKNATVSPISRYP	QY 289	TACATCACTCTCTTGAGAGGGGAGTGAGGGGCTGCTGGCTGTGGCACCAACGCGCCGG 348			
	KRQRTKIMYGPDLRIPNFRSDVMDLLOQOGITVMRVEVGRCLTQDNHELYDIT	DB 3942	TTTATAACTCTGCTTCATTAACCCAGACAGATGGCTGCTAGCTTGTGGTACTAACTACACAG 4001			
	MDMVLKELKSKSQQIYYWEGFENVTEPIHYASDIQVTEGTEYDVDMIMKKIGPD	QY 349	CACCCAGCTGCTGGAACTCTGCTGTAATGCACTGTGGTCCCACTTGGCGAGATGAGAGCG 408			
	VFAQOAVLPNVSIPSTVATALSFKLRAAMYFTKREMSFYAILQCOGNISQISQ	DB 4002	AAACCCAGCTGCTGGCTGATAAACCAACCTTAACAACCTCAATTTTGGGGGCCAAAACCTAGGC 4061			
	TDLGRQKILMNSPLASCLGAYTNFAFCRODYKEMKLYTSLMVOAATNIRAGS	QY 409	TACGCCCCCTTACGCCCCGACGAGAACCTCCCTGGTCTGCTTTTGAAGGGGACGAGGTGTAT 468			
	PTVGGFLLKNNGLKVKLYRPMLEYEHLASSIGLNKLTADFRGNILVALGOQYPS	DB 4062	TTAGCCCCCTTCTACCATCATCTGGCAATCTGGTCTGCTTTTGACCAAGATGACACCTAT 4121			
	VLDTHPFYIRDSAADYKNVQALNGFYCTVTSPCVTSRLDLGKTKVQHLFALLKN	QY 469	TCACCATCCGGAAGCAGGAATAACATGGAAGATCCCTCGGTTCGCGCGCATCCGGGCG 528			
	GCANLYISCLPAALVQLORIPETEMWLSLEHLDDVFENTVANILFLTVKNTSEYTT	DB 4122	TCCACCATTAACCTCTACAAGACCTCAGTGCTCTCACAAAGTTTAGGAGGATCGCTGGC 4181			
	LQOQKPEILLONVARLKCDVHILGEIVDKIGIIVINDLGNPNYNTRRHVGQHCYS	QY 529	GAGAGTGAGCTGTACACCACTGATGCTGTGTCAGAACCCACAGTTTCATCAAGGCCACC 588			
	MFKPNQDITLPGASBEEECVLTPLGLDWAQGVIEECIRNTLLSPVTGSKAF	DB 4182	CAAGTAAAGTAATAACAGAGTGACACCGCCATGCACCGGCCACAGTTTGTCCAGGCAACA 4241			
	VYQHDRCNGTLVQOQGVGRFDPLADYSLVLSLVKPPMGDTDPDSLAMKSLSD	QY 589	ATCGTGCACCAAGACCCAGCTTACGATGACAGATCTACTACTTCTTCGAGAGGACCAAT 548			
	LNINNAEVENHNPASCMAGVQPLKVAASPSNGALRAIAELLNMLGPKMLLSNLI	DB 4242	GCTGTGCATATAAAATGAATCTTATGATGATAAAATCTACTTTTCTTCAAGAAACACG 4301			
	VTASTWESGQCEELKAMLRACRAFAEOLGVNFVSVASSVPDSDRLNTGIPRPS	QY 649	CCTGACAAGAATCCTGAGGCTCTCTCAATGTGTCGCGTGTGCGCCAGTTGTGAGGGGG 708			
	KFEKNVCSGTCKVTSVKRALPVLOKAGNLIHVSDIDPTEGLAGTVFENIYGLSLGV	DB 4302	CACAGTGACTTCAACAGAGTTTCCACATAGTCTACCTAGAGTGGGCGAGGTGCTCTAGT 4361			
	SEISAPKNAALFALIQDQGVGVISGHDSVGLVACVAELCLUSTEFVRVNTVPEGL	QY 709	GACAGGGTGGGAAAGTTTCACTGTGCTCCAAAGTGGAAACACTTTTCTGAAAGCCATG 768			
	HALGYLFAESPVGVEIISPNLPALRDKCSAAGFCODIAYITNGDESFVSVNGKI	DB 4362	GATCAAGGTGGGAGAGCTCCCTGCTCTCAAGTGGACCACTTTTAAAGGCGACA 4421			
	FYQPLSLRATWSTYSGDLSALCSDEEDHRGDYGCNEVDPRGLKEGLMNNYVA	QY 769	CTGGTATGCAAGTGTGCTGCCACCAACAGAACCTTCAACAGGCTGCAACAGCTCTTCCTG 828			
	FYQPDIONKVAIVPWGTKDESFTNAOFTYMCVTEINSPSEFENDFRGLAI	DB 4422	CTGGCGTGTGTAGACTATGATACTGGAAGAATCTACAATGAGCTACAAGATATTTTTCATC 4481			
	GNGVYTPPLVACALATLHTQDGSPLARALLHFNRAOTFSLGCGEFGFLLSQF	QY 829	CTCCCTGACCCCGCCAGTGGAGGGACACCGGCTCTATGGTGTGTTTCTTCCCAACCCC 888			
	IYVLEHQSLFVGHQHEEKLSPISEKTKSGYSERWLSVHPITETGIMLAPCRGMV	DB 4482	TGGCAAGCCCGCAGAGAACAGCTGGGAAGAGACTCTCATCTATGGACTTTTTTGGACCCG 4541			
	LPCWQGYLGLSYKIDGVELINENLVACFSKSSSENFAHRIYPRNPSGNSVA	QY 889	TGGAACCTACTACGCGCTGTGTGTATTCCCTCGGTGACATTGACAAGGTCTTCGCTACC 948			
	GICSDGGRRLATLIDPSLSFPEQWHLPSGYERMOTSPWSLLFYQMHMCCS"	DB 4542	TGGAACCTTTTCTGCGGTCTGTGTGTTACTGTAAAGGACATTGACCATGTGTTTAAAGACA 4601			
	125771..126277	QY 949	TCCTCACTCAAGGGCTACCACCTCAAGCCCTTCCCAACCCGCGGCTGGCAAGTGCCTCCCA 1008			
	/note="similar to Bcl-family of proteins; contains only conserved Bhl domain; functional similarity may exist to ORF16 of H. salmieri, KSHV, BHV4 and EBV BNRF1"	DB 4602	TCCAAAGTAAAAAATATCATCATAACTCCCCACACCTTAGACAGGCAATGTCATGAAG 4661			
	/codon_start=1	QY 1009	GACAGCAGCCCATACCCACAGAGACCTTCCAGGTGGCTGACCGTCAACCCAGAGGTGGCG 1068			
	/evidence-not_experimental	DB 4662	AACCATGACATGTTCCCAACAGAAACCTTTCCAGTTTGGTTCGACAGATATCCAGAAAGTTGCA 4721			
	/product="ORF75"	QY 1069	CAGAGGGTGGAGCCCATGGGGCTCTGAAGAGCGCCATTTCTCACTCTAAATACCACTAC 1128			
	/db_xref="PID:g2338037"	DB 4722	GATCTGTATATACAGAAGAACAAATGCCATGTTTCCAATAATTCAGTCAAAATATATCTAC 4781			
	/translation="WVYITRPDRPPNPPDGAAGCGPCAEVYVANSGGTPDEKGE	QY 1129	CAGAAAGTGGCGTCCACCGCATGCAAGCCAGCCAGGGGAGACCTTTTCATGTGCTTTAC 1188			
	GVKGGGDDPQPLGLDYSNPEKAFGTPASGGUGPIVLPPLPDSDDDESPCPSWY					
	WPLMUSQTEYDKLDRSNEFKDQRRPPTTLALPVCLTTEVLVDGTRWHDNLDG					
	PDSMTPGCGGASPSNGLGAKDKGPKGSPKCGPAGTKTPKDLGARGPKKGP					
	GKSPKDPGAAPKPKDGAPEKPLVLAFFPGPGVDTPNPSHDVTEAAPPDENPKKM					
	WLPPCTKPPCTPLPDCAVLLSSGTPPDWRKGIHRTKVKTIPTVPTVLTISD					
	DDLDRKAGAGGPKKPLPSPRDAGSGQGTTPQSKESTSSGLSGATPRSKAPAGRG					
	ROEPPPLPRDPASPALRLELDIVQKRYIIGVAYNELEAIFRDNRYERERERR					
	MDYALIIILIIIIIVLLIVL"					
	BASE COUNT 35848 a 31711 c 28595 g 34454 t					
	ORIGIN					

Db	4782	ACCAAACTACTTGTATAGAGGTAGAGTATGAGAGTGTATTTTGGGCAACTATTTTAC	4841
Qy	1189	CTAACTACAGACAGGGCCACTATCCACAGGTGTGGAAACGGGGAGCAGGACACACAGC	1248
Db	4842	CTCACTACCAACAAAGGACATTCATATATGTGAGGTATGAAGATCCAACTCTACA	4901
Qy	1249	TTCGGCTTCAACATCATGAGATCCAGCCCTTCCGCCGGGGGTGCCATCCAGACCATG	1308
Db	4902	ACAGCTCTCAACATTTTAGAGATAAATCCCTTTTCAGAAAGCCAGCCCATACAGATATT	4961
Qy	1309	TCGCTGGATGTCGCGGAGGAGCTGTATGTGACCTCCCACTGGAGGTGAGCAGGTG	1368
Db	4962	CTTTTAGATATACAAATCTAAAGCTTTATGTAAATTCAGAGTGGGAGGTGAGTGAGGTG	5021
Qy	1369	CCCTCGGACCTGTGTAGGTCTATGGCGGGGCTGCCACGGTTGCTCATGTCCCGAGAC	1428
Db	5022	CCATTAGACCTATGTTCAGTGTATGGGAATGATGTTTCAGCTGTTTATGTCAAGGAT	5081
Qy	1429	CCCTACTCGGGCTGGGACCAAGCGCTGCATCTCCATCTACAGCTCCGAACGGTCACTG	1488
Db	5082	CCCTGTGCATGTGATAAACAACACCTGTCTCTTAAACAGAGAGTATCTGTTGAAACC	5141
Qy	1489	CTGCAATCCATTAATCCAGCGGACACACAGGAGTGTCCCAACCCCAACACGACAG	1548
Db	5142	GGTGTGTCAGCTTAACCGCACCTTTTCAGAAAT---GTGTGTGACCATATGCTCCAACT	5198
Qy	1549	GGCCCACTGCAGAGGTTTCCTGTGCCCCAACTCTCGCTACTACTGAGTGCCTCCCATG	1608
Db	5199	GTGGTTAGCATCAAGTTCTATACCTCTATATCTATTTCTTATTTCTCTGCCACGA	5258
Qy	1609	GAATCCCGCCAGCCACCTACTCTATGCGGCCCAACAGGAGAACGTGGAGCAGCTCGAA	1668
Db	5259	GTCTCAAAACACCGCTACTACTTTTGGACTAAAGATGTTTCACAGAAAAAAGATGCCAT	5318
Qy	1669	CTGTGTACCAGACCCCAACTGCATCTCTGTTCATCGAGAACCTCACGGCGCAGCAGTAC	1728
Db	5319	GTCAAAACACAAAAATGACTGCATCTGTGCTTATAGCTAACAGCAGCAGCAGCCACTAAT	5378
Qy	1729	GGCCACTACTTCTCGGAGGCCCGCAGGAGGCTC	1760
Db	5379	GGAAACCCAGCTGTGCAACATGAAGAAGATTC	5410
RESULT	9		
LOCUS	AHU18243		
DEFINITION	Alcelaphine herpesvirus 1 putative semaphorin homolog (AHV-sema) and putative membrane antigen genes, complete cds, and major ssDNA-binding protein gene, partial cds.		
ACCESSION	U18243		
NID	61000716		
KEYWORDS			
SOURCE	Alcelaphine herpesvirus 1.		
ORGANISM	Alcelaphine herpesvirus 1.		
REFERENCE	1 (bases 1 to 11874)		
AUTHORS	Ensser, A. and Fleckenstein, B.		
TITLE	Alcelaphine herpesvirus type 1 has a semaphorin-like gene		
JOURNAL	J. Gen. Virol. 76 (Pt 4), 1063-1067 (1995)		
MEDLINE	97201573		
REFERENCE	2 (bases 1 to 11874)		
AUTHORS	Ensser, A.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-DEC-1994) Armin Ensser, Institut fuer Virologie, Friedrich-Alexander-Universitaet Erlangen-Nuernberg, Schlossgarten 4, Nuremberg, Bavaria 91054, Germany		
FEATURES	Location/Qualifiers		
source	1. .11874		
	/organism="Alcelaphine herpesvirus 1"		
	/strain="C500"		
	/specific_host="Connochaetes taurinus taurinus"		

/db_xref="taxon:35252"			
/lab_host="Bos taurus"			
/clones="clones C238, C249, C500BS"			
<1. .(330.346)			
/notes="gamma-2 herpesviral GC-rich terminal tandem repeats"			
/function="putative packaging signals"			
/rpt_type=tandem			
3573. .5534			
/standard_name="AHV-sema"			
/notes="putative semaphorin homolog; similar to Genbank Accession Number L26081"			
/codon_start=1			
/db_xref="PID:g1000717"			
/translation="MAYLNATVSKPVISLSKVKLFKFGHCGGCGCGLITEFVIH PAMGTLGVISRLMILSAITAKSRFIDKPLRVLNLTDFGQGHRRFGPQEPVTLVLFH SLNSDDVYGGNTIYLFDFHSSNASTALINITSTHNRLSSTCENTETLLHNOD GLLACSTNQSKSCHLINLITOFGLPKGLAFSPSSGNLVLFDDNDYISTINLKS LSGSHFRRIAGVELYTSDTAMHRPOFOVATAVHKNESYDDKIYFFQFNSHSDKQ FPHYPRVQGVCSDDQGSSESVYKWTTLFKARLACVDYDTGRIINELQDIFWQAP ENSWEETLIVGLFLSPWNSAVCVFTVKDIDHVKTSKLNHYHKKIYPRPGCKMNH QHVPTTFQVADRYPEVADPVYQKNAMFPIQSKYITKLLVRYVEYGVVATIFY LTIKGTIHYRYEDNSNTALNILEINFQKPAQIONILLDNTNLLKYVNSEWYS EVLDCSYGNDGFCFSCMRDPLTWNITKTSFKORVSETGGTPANRILNEMCGDHY APTVAHQVSIPLLSNYSILSCPASNHDYFTWKDGTETKRCHVKTHKNDICILLIANS TTATNGTHVCNMKDSVTYKLVLEVNTLM"			
6414. .10523			
/notes="putative membrane antigen; similar to p140 (PIR Accession Number S29606), and p160 (Genbank Accession Number Z26584)"			
/codon_start=1			
/db_xref="PID:g1000718"			
/translation="MITRELQLLAVTAEPSALETAIAALRSIPQLQNVITOTDAYL VTFSTPRLREHQLAETLFTKAAALRTFDEQVLPPLPSALSSTFTVYGGDINRL PFSQNELTALQTRDAREKVENVERIQHCRLLFFSGPGEQLNHHYHALLRILGCD LTIQPRNELNSSLDFATYTPPRLADKRVHYVHGDIINLITGAASNYNWSPT AVTDSIFRIPYTHVVRVDRWEIRVRVPDVGFGHGRVFRQHGGAAYCMYCNLTST GATGCLQKMITSFISQPTLQGLITPEMCGWQOELSATIOMLTKYANTVEAFK TLNAGVPLVQGVFTLSPTLREKRAMAFTGSLITGLPPVILKPPTAETLDAAMERNR QALLVEGYDPYDGKGLHNPVIMNSESGRHAHLHQAOLRLMIITPGKVIHICCDW QYNEKTMFKIAKACGDLGSLMSARSLPHTLRLKKWFRNLNHYHRIKNSWLKVD SAAALVIADNEYQDVSKRMDIALSGWCPFHILGNLTIPNSNTIVISDNQYGEIVDI QYKMHMPKQSEGTEDTLPAPMSNLQFLKNDLDELQLLQVLRHPTVCGKAHIVHV DCGNGHIAQQGVGPFDPCLDVSFVHNLDVGDIREGSESVRYMAADRYVARRLI ETQYSTGLDITDATTANLGLTYHIPSEHOVTVESKYGNCIGEKITFTQRPDL GTLTAIVESCNCILGVNREFLIGLSIVPEGIHYRHEVNSIMAKADFCSSNF GQVNSAENGNCILRSVATANAPCVVPGPSLKYFKPGSAILRNLHTEFLSGGI CCMASIGASEFTPTPQSLRNLQFLNVLKAEALNSGDHSDGGLICACEMMFAG GLSARLIIHDEDEEPVPLFSETPGFVLENAIDVAAIIRANLYNVECIQGEVVEVS DTFTVFHNTQLLSVPVRLKHNWTLFSKVDLLYVKEDQVLPETSYGNYEVHLLTV PYISQSTTRPNVLVHLLPGGYPDALLAALTNSGSPDTVYVPGCKYRHNREDAA PGSPTADDTAGIVLYSSNIDSDVDGSTIROWLNVRQVINDVRNLAKGSETLAI GOLACRILEATKAIGFDAGSOOTPELLPNASRYSEWLNFKIPEDTKAVAFRLAGC VLPWVQGHLLGFSHNNTIFFGDLERQQVAATFNGLPVQSGPAREYPLNPTBAEHPY AGLCSDEGRHLALDFCLAFNTQWQHNQTPGQGLPVPSPWKLFRYLRYLNWSKFHQ HYRSLRLNLRHTFNFETQQLDFPHADRRAVPHDNPQYIPMDAMPDPSQ"			
11341. .>11874			
/codon_start=1			
/product="major ssDNA-binding protein"			
/db_xref="PID:g1000719"			
/translation="MALKHQLNHAVEDNLGSKRAPIGCPGFIYIYPETHFNFKRISLLG DRYAEAGAFSLPLLHGVTVEAFVPNVKAIVKIDMTVTSVKLSITFNRAIFHNVEK FESIFSGPLGSLCKEACDLFGYVPTPLGEGSTDVSDICPPVWQEKDAIMAVITEG EKERHLGLIYLLKSOH"			
BASE COUNT	3437 a	2699 c	2362 g 3376 t
ORIGIN			
Query Match	20.0%;	Score 380;	DB 17; Length 11874;
Best Local Similarity	52.3%;	Pred. No. 1.2e-67;	
Matches	864;	Conservative 0;	Mismatches 785; Indels 3; Gaps 1;
Qy	109	TTTGGCCAGACTGAGCGCACACGGTGCTTTTCCACGAGCAGCAGCTCCTCTGTGTGG	168


```

/gene="SEMAL"
/product="semaphorin L"
/evidence=experimental
<1. >8265
/gene="SEMAL"
join(<1..114,605..646,944..1036,1238..1322,1443..1553,
2176..2315,2907..3091,3945..4053,4146..4344,6876..7003,
7178..7332,7533..7594,7904..8265)
/gene="SEMAL"
/note="sequence starts with the putative second exon of
SEMAL; H-SEMA-L"
/codon_start=1
/product="semaphorin L"
/db_xref="PID:g3551779"
/translation="EPHTLVFPEGSSVWVGGRGVYLFDFPECKNASRVTVNIGST
KGCGLDRKDCENYITLLERRSEGILACGTNARHPCWNLVNGTVVPLCEMGYAFSP
DENSELVFEGDEVSTYRKQYNGKIPFRKIRGESELYSTDVWQNFQKATIVHQ
DQAYDKKIYFFEDNDNPAPLNVSRVAQLCGDGGESSLVSXWNTFLKAMLY
CSDAATNKNRNLQDVLFPDQGWDRTRYGVFSNPWTSAVCVSLGIDIKVFRT
SSLKGYHSNLRPFGKLPQQPIPTETFOADRHPEVAQRVEPMGPKLPFLHSH
IYQVAVHMQAASHGTFHVLVLTTRDGTIKHVVEPGEQHSFAFNIEMIOFPFRAAA
HQTWLSDAERBKLXVSOWSEYVSQVPLDCEYVGGCGHCLMSRDPYCGWDGRCLSIY
SSERSVQSINAPBPHRECNPDKAPLQKVSAPNSRXYLSCPMNSRHGRTISWRHK
ENVSQCEPGHQSNGLFIENLTAQOYGHYFCCAQEGSYFRAQHWOLLPEDGTMAE
HLLHGACALAAASLWGLVPLTGLLHV"
1749 a 2358 c 2757 g 1997 t

```

Query Match	19.2%	Score 366;	DB 12;	Length 8861;
Best Local Similarity	96.2%;	Pred. No. 8.5e-65;		
Matches 375;	Conservative 0;	Mismatches 15;	Indels 0;	Gaps 0;
Qy 1513	CCACACAGGAGTGTGCCAACCCCAACCCAGACAAGGCCACCTCGAAGAGTTTCCCTG	1572		
Db				
Qy 1573	CCCCAAACTCTCGCTACTACTGAGCTGCCCATGGAATCCGCCACGCCACCTACTCA	1632		
Db				
Qy 1633	TGGGCCACAAAGGAGAAGCTGGACAGAGCTGCCGAACCTGGTCAACGAGAGCCCAACTGC	1692		
Db				
Qy 1693	ATCTGTTTCATCGAAGACTTCACGCGCAGACAGTACGGCCACTACTTCTGGAGAGCCAC	1752		
Db				
Qy 1753	GAGGCTCCTACTTCCGGAGGCTCAGACATGGCAGCTGTGCCGAGAGCGGCATCATG	1812		
Db				
Qy 1813	CCCAGACACCTGCTGGGGTCATGCCCTGGCCGCGCTCCCTCTGCGTGGGGGTGCTG	1872		
Db				
Qy 1873	CCCACATCACTCTTGGCTTGTGCTGCAC	1902		
Db				
Qy 8233	CCCACATCACTCTTGGCTTGTGCTGCAC	8262		
Db				

RESULT	11
AF030697	
LOCUS	AF030697 8861 bp DNA PRI 09-SEP-1998
DEFINITION	Homo sapiens semaphorin L (SEMAL) gene, partial cds.
ACCESSION	AF030697
NID	g3551778
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
	Primates; Catarrhini; Hominoidea; Homo.

REFERENCE	1 (bases 1 to 8861)
AUTHORS	Lange,C., Liehr,T., Goen,M., Gebhart,E., Fleckenstein,B. and Ensser,A.
TITLE	New eukaryotic semaphorins with close homology to semaphorins of DNA viruses
JOURNAL	Genomics 51 (3), 340-350 (1998)
MEDLINE	98389619
REFERENCE	2 (bases 1 to 8861)
AUTHORS	Ensser,A., Liehr,T., Lange,C., Goen,M., Gebhart,E. and Fleckenstein,B.
TITLE	Direct Submission
JOURNAL	Submitted (22-OCT-1997) Institut fuer Klinische und Molekulare Virologie, Friedrich Alexander Universitaet, Schlossgarten 4, Erlangen, Bavaria 91054, Germany
FEATURES	Location/Qualifiers
source	1. .8861
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/map="15p22.3-p23"
	/chromosome="15"
mRNA	join(<1..114,605..646,944..1036,1238..1322,1443..1553,2176..2315,2907..3091,3945..4053,4146..4344,6876..7003,7178..7332,7533..7594,7904..>8265)
	/gene="SEMAP"
	/product="semaphorin L"
	/evidence=experimental
gene	<1..>8265
	/gene="SEMAP"
CDS	join(<1..114,605..646,944..1036,1238..1322,1443..1553,2176..2315,2907..3091,3945..4053,4146..4344,6876..7003,7178..7332,7533..7594,7904..8265)
	/gene="SEMAP"
	/note="sequence starts with the putative second exon of SEMAP: H-SEMA-L"
	/codon_start=1
	/product="semaphorin L"
	/db_xref="PID:g3511779"
	/translation="EPHTVLFPEHSGSSVWVGGRKVYLFDPPEKKNASRVTVNGIST KGSCLDKDCENYITLLRSEGLKACGTNARHPSCWNLVNGTVVPLGEMRGYAPFSP DENSLVPEGDEVSTYTKQBYNGKIPRFRIGRSELYTSDTVNQNFQIKATIVHQ DQAYDKIYFFREDNDKPNPELNSRVAQLCGDGOGGESSLSKWNTEFLKAWLV CSDAATNKNRNLQDVELLPDPGOWDRITRVYGSNPNYSACVYSLGDDKVFRT SLLKGYHSSLPNRPQDPQPIPTETFOVADRHPEVAQVEPMGLTKPLPFHSKY HYQKVAVRWQAGHGFHVLVLTDRGTIHKVVEPGEQHSFAENIEIQFPRAAA IQTMSLDAERKLVYSQWSEYVPLDCEVYGGCGHCLMSRDPYCGWDGRCISY SSERSVLSQSNPAPHEPCPNPKDKAPLQKVSAPNSRYLSCSPESRHATYSRHHK ENFQSCSEQSCILFIENLTCAQOGHYFCAQEGSYFREAQHWOLLPEDGIMAE HLHGACALASLWGLVLPITLGLLVH"
BASE COUNT	1749 a 2358 c 2757 g 1997 t
ORIGIN	

Query Match	19.2%	Score 366;	DB 42;	Length 8861;
Best Local Similarity	96.2%;	Pred. No. 8.5e-65;		
Matches 375; Conservative	0;	Mismatches 15;	Indels 0;	Gaps 0;
Qy 1513	CCACACAGGAGTGTCCCAACCCCAACACGACAGAGGCCCACTGCAGAGAGGTTTCCTCG	1572		
Db 7873	CCTACGGCTGTCTCCCACTGCTCCCCACAGACAGGCCCACTGCAGAGAGTTTCCTCG	7932		
Qy 1573	GCCCCAAACTCTCGCTACTTACTGTAGCTGCCCATGGAATCCGGCCACGCCACTACTCA	1632		
Db 7933	GCCCCAAACTCTCGCTACTTACTGTAGCTGCCCATGGAATCCGGCCACGCCACTACTCA	7992		
Qy 1633	TGGCGCCACAAAGAGACGTGGAGCAGAGCTGCGAACTGGTCCACAGAGCCCCCACTGC	1692		
Db 7993	TGGCGCCACAAAGAGACGTGGAGCAGAGCTGCGAACTGGTCCACAGAGCCCCCACTGC	8052		
Qy 1693	ATCTGTTCATCGAGAACTCAACGGGGCAGCAGTACGGGCACACTTCTCGAGGCCACG	1752		
Db 8053	ATCTGTTCATCGAGAACTCAACGGGGCAGCAGTACGGGCACACTTCTCGAGGCCACG	8112		
Qy 1753	GAGGGCTCCCTACTTCGCGCAGGCTTCAGCACTGGCAGCTGCTGCCCGAGACGCAATCATG	1812		

```
|||||
Db 8113 GAGGCTCTACTTCCGCGAGGCTCAGCACTGGCAGCTGCTCCCGAGGACGGCATCATG 8172
Qy 1813 GCCGAGCACCCTGCTGGGTCATGCGCTGTGCCCTGCGCCGCTCCCTCTGCTGGGGTGTG 1872
Db 8173 GCCGAGCACCCTGCTGGGTCATGCGCTGTGCCCTGCGCCGCTCCCTCTGCTGGGGTGTG 8232
Qy 1873 CCCACACTCACTCTTGGCTTGGCTGCTGCAC 1902
Db 8233 CCCACACTCACTCTTGGCTTGGCTGCTGCAC 8262

RESULT 12
HSU38276 3552 bp mRNA PRI 05-NOV-1996
LOCUS Human semaphorin III family homolog mRNA, complete cds.
DEFINITION U38276
ACCESSION 91061350
NID
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 3552)
AUTHORS Xiang,R.H., Hensel,C.H., Garcia,D.K., Carlson,H.C., Kok,K.,
Daly,M.C., Kerbacher,K., van den Berg,A., Veldhuis,P., Buys,C.H.
and Naylor,S.L.
TITLE Isolation of the human semaphorin III/F gene (SEMA3F) at chromosome
3p21, a region deleted in lung cancer
JOURNAL Genomics 32 (1), 39-48 (1996)
MEDLINE 96230324
REFERENCE 2 (bases 1 to 3552)
AUTHORS Naylor,S.L.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1995) Department of Cellular & Structural
Biology, University of Texas Health Science Center, 7703 Floyd Curl
Dr., San Antonio, TX 78284-7762, USA
LOCATION/Qualifiers
FEATURES
source
1..3552
/organism="Homo sapiens"
/notes="Sequence confirmed by RT-PCR"
/db_xref="taxon:9606"
/clone="N5, N10, N12"
/clone_lib="Stratagene fetal brain cDNA in lambda ZAPII
(Cat 936206)"
/map="3p21.3"
/chromosome="3"
/sex="female"
/tissue_type="brain"
/dev_stage="fetal (17-18 weeks)"
1..315
316..2577
/codon_start=1
/product="semaphorin III family homolog"
/db_xref="PID:g1061351"
/translation="MLVAGLLWASLLTGAWPSEPTQDHLPATPRVRLSKFKELKATGT
AHFNLNTDTRILLKDEDDHDMVSGKDYVLSLDLHDINREPLIIHWAASQRIE
EVLGSKDNGECGFVLPQWNRTHLYVCGTGAYNPMCTYVNRGAAQDYIFYLEP
ERLESKGKCPYDKLDTASALINEELYAGVYIDFMGTDAAFRTLGKQAMRTDQYN
SRWLNDSFTFHAELIPDSAEKDKLVFFFRSAGAPSPAVYARIGRLCLNDGGHC
CLYNKSTFLKARLVCSVPGEDGIEHFDLQVFOOTODVRNPVIYAVFTSSGVF
RGSACVYSNADIRNMFNGPFAKEGPNQWMPFSKMPYPRGTCGGTFTSMKST
KDPDEVINMRSHPLMYQAVYPLQRLPVRTGAPYRLTIAVDQVDSADGRIEVL
LGDRGTQVIVLPKDDQEMELMLEEVEVFDPAPVKMTISSRQQLYVASAVG
THLSLRQYAGAACDCCLARDPYCAWDQAGSRYSASSRRSRQDVRHGNPIRQC
RGFNSNANKNAVESVOYGVAGSAFLPCQSPQATVKWLFQDPGDRREIRAEDEF
LRTEOGLLRALQSDRLGYSCATENNFKHVTVOLHVLGRDVAHAALFPLMSA
PPPGACPPPTPQELAQLLAQVEGLIHQYCOGYWRHVPSPREAPGAPSPQDQ
KKPRNRHHPPPTD
2574..3546
736 a 1100 c 991 g 725 t

BASE COUNT
ORIGIN
```

```
Query Match 3.7% Score 70.6; DB 10; Length 3552;
Best Local Similarity 55.4%; Pred. No. 4e-05;
Matches 158; Conservative 0; Mismatches 124; Indels 3; Gaps 1;

Qy 1164 CGGGGAGACCTTTTCATGTGCTTTTACCTAACTACAGACAGGGGCGACTATCCACAAGG---T 1220
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1641 CGAGCGGCGTATGAGGTGCTTTTCTGGGCACAGACGCGGACAGTGCAGAGGTGAT 1700

Qy 1221 GGTGGAACCGGGGAGGAGGACAGCAGCTTCGCTTCAACATCATGGAGATCCAGCCCTT 1280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1701 TGTGCTGCCCAAGGATGACCAGGAGATGGAGGAGCTATGCTGGAGGAGGTGGAGGTCTT 1760

Qy 1281 CCGCCGCGCGCTGCCATCCAGACCATGTCGCTGGATGCTCAGCGGAGGAGCTGTATGT 1340
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1761 CAAGATCCAGACCCCGTCAAGACCATGACCATCTCTTCAAGAGGCAACAACCTCTACGT 1820

Qy 1341 GAGCTCCAGTGGGAGGTGAGCCAGGTGCCCTGAGCTGTGTGAGGTCTATGCGGGGG 1400
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1821 GCGCTACGCGGTGGGTGTACACACACTGAGCCTGCACCGCTGCCAGCGGTATGCGGGTGC 1880

Qy 1401 CTGCCAGGTTGCTCATGTCCCGAGACCCCTACTCGGCTGGGA 1445
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1881 CTGTGCTGACTGCTGCTTGCCTGCGCGGACCTTACTGTGCTGGGA 1925

RESULT 13
HSU33920 2719 bp mRNA PRI 10-JUL-1996
LOCUS Human clone lambda 5 semaphorin mRNA, complete cds.
DEFINITION U33920
ACCESSION 91000206
NID
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2719)
AUTHORS Roche,J., Boldog,F., Robinson,M., Robinson,L., Varella-Garcia,L.,
Swanton,M., Waggoner,B., Fishel,R., Franklin,W., Gemmill,R. and
Drabkin,H.
TITLE Distinct 3p21.3 deletions in lung cancer and identification of a
new human semaphorin
JOURNAL Oncogene 12 (6), 1289-1297 (1996)
MEDLINE 96226360
REFERENCE 2 (bases 1 to 2719)
AUTHORS Drabkin,H.A.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1995) Harry A. Drabkin, Division of Medical
Oncology, University of Colorado Health Sciences Center, 4200 East
9th Ave., Denver, CO 80262, USA
LOCATION/Qualifiers
source
1..2719
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="lambda 5"
/clone_lib="Stragene fetal brain"
/chromosome="3"
/map="3p21.3"
/tissue_type="brain"
/dev_stage="fetus"
79..2436
/codon_start=1
/product="semaphorin"
/db_xref="PID:g1000207"
/translation="MLVAGLLWASLLTGAWPSEPTQDHLPATPRVRLSKFKELKATCT
AHFNLNTDTRILLKDEDDHDMVSGKDYVLSLDLHDINREPLIIHWAASQRIE
EVLGSKDNGECGFVLPQWNRTHLYVCGTGAYNPMCTYVNRGAAQATPTQTQ
AVRGSRATDGAALRPMPTAPRODYIFYLEPERLESKGKCPYDKLDTASALINEEL
YAGVYIDFMGTDAAFRTLGKQAMRTDQYNRSLNDPSFHAELIPDSAEKDKLV
FFFRSAGAPSPAVYARIGRLCLNDGGHCLLVNKSFTLKARLVCSVPGEDGIEH
HEFDLQVFOOTODVRNPVIYAVFTSSGSVFRGSACVYSNADIRNMFNGPFAKREG
PNQWMPFSKMPYPRGTCGGTFTFSMKSTKDYIDPEINFRSHPLMTQAVIPLQR
```


RPLWRTGAPRULTTIAVOQDVADGDRYEVLFGTRDGTQVKVIVLPPKDDQEMEELML
EEYEVKSPATVMTISSRQOGLYVASGVTHLSLHRQOYGAACACCLARDPYC
AWDQCSRTASSKRSQDHYANIGRIQCRGENSNANKNAVESVOYGVASAALF
ECQPSQATVKWLFQDPCDGRREIPAEDRFLRTEOGLLRALQLSDRGLYSCATATE
NNEKHVVYRVOLHVHVAALFPLSPASPPPGAGPPTPYQELAQLLAQPEVYG
LH1QYCOGYWRHVPSGREAPADGAPPEQDOKKPNRRRHPPDT "

BASE COUNT	551 a	872 c	768 g	528 t
ORIGIN	LBNQACGQIMNNVFFAFANAFGRFASFEPVQQAANKKRRHFFDI			

Query Match	3.7%	Score 70.6;	DB 12;	Length 2719;
Best Local Similarity	55.4%	Pred. No. 4.2e-05;		
Matches 158;	Conservative	0;	Mismatches 124;	Indels 3;
				Gaps 1;

[illegible]

RESULT	14	
HSU38276		
LOCUS	3552 bp	mRNA
DEFINITION	Human semaphorin III family homolog	mRNA
ACCESSION	U38276	
NID	g1061350	
KEYWORDS	.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;	
	Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	

REFERENCE
AUTHORS
1. Zetterstrom, R.H., Hensel, C.H., Garcia, D.K., Carlson, H.C., Kok, K., Xiang, R.H., Hensel, C.H., Kerpacher, K., van den Berg, A., Veidhuls, P., Buys, C.H. D.M., and Naylor, S.L.
TITLE
Isolation of the human semaphorin III/F gene (SEMA3F) at chromosome 3p21, a region deleted in lung cancer
JOURNAL
Genomics 32 (1), 39-48 (1996)

MEDLINE
 86203028 32 (17, 35-46 (1990))
 REFERENCE
 2 (bases 1 to 3552)
 8230324
 AUTHORS
 Naylor, S.L.
 TITLE
 Direct Submission
 Submitted (11-OCT-1995, Department of Cellular & Structural
 Biology, University of Texas Health Science Center, 7703 Floyd Curl
 Dr., San Antonio, TX 78284-7762, USA)

```

FEATURES
source
    1. .3552
    Location/Qualifiers
        /organism="Homo sapiens"
        /note="Sequence confirmed by RT-PCR"
        /db_xref="taxon:9606"
        /clone="N5, N10, N12"
        /clone_lib="Stratagene fetal brain cDNA in lambda ZAPII
            (Cat 936206)"
        /map="3p21.3"
        /chromosome="3"
        /sex="female"
        /tissue_type="brain"
        /dev_stage="fetal (17-18 weeks)"

```

5' UTR
CDS

```

316. .25//
/codon_start=1
/product="semaphorin III family homolog"
/db_xref="PID:gl016351"
/translation="MLVAGLLLDWASVLTGAGSFFQDHLPATPRVRLSRFKELKATGT
EVLSSGKDVGCGENFVRIQLQPNRTHLYVCGTGAYNCTQYVNRGRRADQYIFYLEP
AHFFNLLNTDYLRIKDDHEDHWSYSGKDYVLSLDLHDINREPLIIHWAASQRTIE
ECVLSGKDCPCYDKLDTIANEELYAGVYIDFMGTDAALFTGLKQAMRTDQVYN
SWLNDSPPIHAELIPDSANEDKLVEFFRESAEAPQAPVAVYRIGLCLNDGGHCL
CLVKNWSTLKLKVCSPGEGDIEHFDLQDVFVQQQDQDVENPIYAVFTSSGSVFI
RSGVACYSMAIDRMVFNFGPAHKEGPNYQWMPFSKGMPTPREGTCPGFTFTFSMTSI
KQYDDEVINRMSHPYQVAPVPLQRRLVYVTRGAPYRLTYIAVDQSDADRGREVLFI
LGTDRVQKIVLPKDDQEMLEELVEEYFKDPAKPTMTIISRRQQYLYVASAVG
THLSLRNQAAYGAACDCLACAPYCAWPGQACSRYSATSKRSRQDVRHNGPIROCI
RGFNHNKNAQYVQYVAGSAAFLCECQPSRQATYKWLFDQPDGRREIRAEFRDI
LRTQGLLRALQLSDRGYLSCTATENNFKHVYRVQLHVLGRDAVHAALFPLLSMA
PPYGGAGPPPTPYQELAQLLAAPVEGLLIHQYCGQYWRHVPPSPREAPGAPSPPEPQD
KKPRNRNRHHPDPT"
2574. .3546
3' UTR

```

3'UTR	BASE COUNT	ORIGIN
ATGTAAGGATGCTT	736 a	991 g
2574...3546	735 t	

Query Match	3.7%	Score 70.6;	DB 40;	Length 3552;
Best Local Similarity	55.4%	Pred. No. 4e-05;		
Matches 158:	Conservative	0. Mismatches	124.	Indels
				Gaps
				1:

QY	1164	CGGGGAGACCTTTTCATGTGTGCTTTACCTTAACATACAGACAGGGGCACACTATCCACAAGG---T	1220
DB	1641	CGACGGGCGGTATAGGTGTCTTTTCTGGGCACAGACCGGGGACAGTGCAGAAAGGTCAT	1700
QY	1221	GGTGAACCGGGGAGCAGGAGCACAGCTTCGGCTTCAACATCATGGAGATCCAGGCCCTT	1280
DB	1701	TGTGCTGCCCAAGATGACCGAGAGATGGAGGAGCTCATGCTGGAGGAGGTGGAGGTCTT	1760
QY	1281	CGCCGCGGGGCTGCCATCCAGACCATGTGCGTGGATGCTGAGCGGAGGAAGCTGTATGT	1340
DB	1761	CAAGGATCCAGCACCGGCTAAGACCATGACCATCTCTTCTAAGAGGCAACAACCTCTACGT	1820
QY	1341	GACCTCCCACTGGGAGGTGAGCAGGTGCCCTTGACCTGTGTGAGGTCTATGCGGGGG	1400
DB	1821	GGCGTCAGCGCTGGGTGTACACACCTGACCTGCACCGCTGCCAGCGGTATGGGGCTGC	1880
QY	1401	CTGCCACGGTTGCTCATGTGCCGAGACCCCTACTGCGGTGGGA	1445
DB	1881	CTGTGCTGACTGTGCTCTGCCGGGACCCCTACTGTGCTGGGA	1925

RESULT 15

RESULTS	RESOLUTION	LOCUS	HSU33920	2719 bp	mrna	PRI	10-JUL-1986
		DEFINITION	Human clone	lambda 5	semaphorin mRNA,	complete cds.	

KEYWORDS

human.
SOURCE
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2719)

1 (pages 1 to 2123)	Roche, J., Boldog, F., Robinson, M., Robinson, L., Varela-Garcia, L., Swanton, M., Waggoner, B., Fishel, R., Franklin, W., Gemmill, R. and Drabkin, H.
REFERENCE	Distinct 3p21.3 deletions in lung cancer and identification of a
AUTHORS	
TITLE	

JOURNAL
Oncogene 12 (6), 1289-1297 (1996)
MEDLINE
96226360

REFERENCE
20220000
2 (bases 1 to 2719)
Drabkin, H.A.
Direct Submission
TITLE
Submitted (14-AUG-1995)
HARRY A. DRABKIN, Division of Medical
Oncology, University of Colorado
Health Sciences Center, 4200 East

JOURNAL
IN RE
DIRECT SUBMISSION
Submitted (14-AUG-1995) Harry A. Drabkin, Division of Medical
Oncology, University of Colorado Health Sciences Center, 4200 East

9th Ave., Denver, CO 80262, USA

FEATURES

Source
1. 2719
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="lambda 5"
/clone_lib="Stragene fetal brain"
/chromosome="3"
/map="3p21.3"
/tissue_type="brain"
/dev_stage="fetus"
79. .2436
CDS
/codon_start=1
/product="semaphorin"
/db_xref="PID:q100207"
/translation="MLVAGLLNLGSLTGWSPPTQDHLDPATPRVLSFKELKATGT
AHFENLLNTDYRLIKDEHDHMYGSKDYLSLDLHDINREPLIIHWAASPORIE
ECVLSGKDVNGECNFEVRLIQWNRTHLYVCGTYNPMCTYVNRRAQAATPWTQTQ
AVRGGRSRTGALRPMPTAPQDYIFVLEPERLESQKCPYDKLDTASALINEEL
YAGVYIDFMGTDAIIFRTLGKOTAMRTDOYNSRWLNDPSFIHAELIPDSAEKNDKLY
FFFRBSAEPQSPAVVARIIGRICLNDGHCCLVKNKSTFELKARLYCSVPGEDGIET
HFELQDVFEVQTDQVNRNPYIAYVTTSSGVSRGSAVCVISMADIRMVNFNGFFAHKEG
PNYQWMPFSRPMPTPGTCGFTPTSMKSTKDYDPEVINEMRSHPLMQAVYPLQR
RPLVYRTGAPVRLTIAVDQDAGDGRYEVLFGLTDRGTQKIVLPKDDQEMEELML
EEVEVFXDPAPVKTMITISSKQQLYVASAVGTHLSLHRCQAYGAACACCLARDPYC
AMDGACSRVYASSKRRSRRODRHGNPIROCGRFNSNANKNAVESQYGVAGSAFL
ECOPRSQATVKWLFORDPGDRREIRAEIDREFLTEOGLLRALQLSDRGLYSCATYE
NNFKHVTIRVOLHVLGRDAVHAALFPLSLNSAPPPGAGPPTFYIQLAQLLAQPEVG
LIHOYCGIWRHVPPSPREAPGAPRSPEDQKKRNRHHPPDT"

BASE COUNT 551 a 872 c 768 g 528 t
ORIGIN

Query Match 3.7%; Score 70.6; DB 42; Length 2719;
Best Local Similarity 55.4%; Pred. No. 4.2e-05;
Matches 158; Conservative 0; Mismatches 124; Indels 3; Gaps 1;
Qy 1164 CGGGGAGACCTTTTCATGTGCTTTTACCTTAACACAGAGGGGACCTATCCACAAGG---T 1220
Db 1500 CGACGGGGCTATGAGGTGCTTTCTGGGCACAGACCGGGGACAGTGCAGAAGGTCA 1559
Qy 1221 GGTGAACCGGGGAGCAGGACACAGCTTCGCCCTTCAACATCATGGAGATCCAGCCCTT 1280
Db 1560 TGTGCTGCCCAAGGATGACAGGAGATGAGGAGCTCATGCTGGAGAGGTGGAGTCTT 1619
Qy 1281 CCGCCCGGGGGTGCATGCCACACCATGTCGCTGGATGCTGAGCGGAGGAAGCTGTATGT 1340
Db 1620 CAAGGATCCAGCACCCGTCAGACCATGACCATCTCTTTTAAGAGGCAACACTCTACGT 1679
Qy 1341 GAGCTCCAGTGGAGGTGAGCGAGGTGCCCTGGACCTGTGTGAGGTCTATGGCGGGG 1400
Db 1680 GCGGTGAGCGGTGGGTGTACACACCTGAGCCTGCACCGCTGCCAGGCGTATGGGGCTGC 1739
Qy 1401 CTGCCACGGTTCGCTCATGTCCCGAGACCCCTACTCGCGCTGGGA 1445
Db 1740 CTGTGCTGACTGCTGCTTGGCCGGGACCCCTTACTGTGCTGGGA 1784

Search completed: July 8, 1999, 19:45:22
Job time: 3481 sec

THIS PAGE BLANK (USPTO)

14
WIREH

(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:39:59 1999; MasPar time 7.88 Seconds
Tabular output not generated. 62.337 Million cell updates/sec

Title: >US-09-041-236-2
Description: (237-245) from US09041236.pep (17 of 45)
Perfect Score: 53
Sequence: 1 DQGGESSLS 9

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 19.684; Variance 19.490; scale 1.010

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	53	100.0	393 11	SEMAPHORIN L (FRAGMENT)	7.83e-03
2	53	100.0	653 14	SIMILAR TO GENBANK ACC	7.83e-03
3	53	100.0	666 4	SEMAPHORIN L	7.83e-03
4	44	83.0	1032 5	MAJOR TEGUMENTAL ANTIG	2.30e+00
5	42	79.2	296 5	F10G8.7 PROTEIN.	7.44e+00
6	42	79.2	314 14	ORF25 PROTEIN.	7.44e+00
7	42	79.2	356 7	MHC CLASS I AOTR-G*03	7.44e+00
8	42	79.2	357 7	MHC CLASS I AOTR-G*01	7.44e+00
9	42	79.2	402 5	F39B2.10 PROTEIN.	7.44e+00
10	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
11	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
12	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
13	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
14	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
15	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
16	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
17	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
18	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
19	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
20	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01

21	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
22	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
23	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
24	40	75.5	150 14	PHOSPHOPROTEIN (FRAGME	2.31e+01
25	40	75.5	187 10	TRANSCRIPTION FACTOR (2.31e+01
26	40	75.5	270 14	(NJ) NS PROTEIN (WITH	2.31e+01
27	40	75.5	274 14	(STRAIN 10/82-CR-B) PH	2.31e+01
28	40	75.5	274 14	(STRAIN 09/82-HD-B) PH	2.31e+01
29	40	75.5	274 14	(STRAIN . /49-UT-B1)	2.31e+01
30	40	75.5	274 14	(STRAIN . /76-EC-M)	2.31e+01
31	40	75.5	274 14	(STRAIN 10/84-GM-P) PH	2.31e+01
32	40	75.5	274 14	(STRAIN 07/84-OA-B) PH	2.31e+01
33	40	75.5	274 14	(STRAIN . /60-PN-B) P	2.31e+01
34	40	75.5	274 14	(STRAIN 11/84-HD-B1) P	2.31e+01
35	40	75.5	274 14	(STRAIN 01/84-SN-P1) P	2.31e+01
36	40	75.5	274 14	(STRAIN 11/82-VC-B2) P	2.31e+01
37	40	75.5	274 14	(STRAIN 07/83-NC-P) PH	2.31e+01
38	40	75.5	274 14	(STRAIN 10/85-HD-B1) P	2.31e+01
39	40	75.5	274 14	(STRAIN 01/85-PN-B1) P	2.31e+01
40	40	75.5	565 10	CARROT B2 PROTEIN-LIKE	2.31e+01
41	40	75.5	588 2	PGRS-FAMILY PROTEIN.	2.31e+01
42	40	75.5	653 4	KIAA0516 PROTEIN (FRAG	2.31e+01
43	40	75.5	701 2	EXO-POLYGALACTURONASE	2.31e+01
44	40	75.5	744 3	NADH DEHYDROGENASE SUB	2.31e+01
45	40	75.5	1380 5	W04A4.1.	2.31e+01

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	393 AA.
ID O88371			
AC O88371;			
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)			
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DE SEMAPHORIN L (FRAGMENT).			
GN SEMAL.			
OS MUS MUSCULUS (MOUSE).			
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC SCUTIROGNATHI; MURIDAE; MURINAE; MUS.			
RN (1)			
RP SEQUENCE FROM N.A.			
RX MEDLINE; 98389619.			
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;			
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA			
RT viruses."			
RL GENOMICS 51:340-350(1998).			
DR EMBL; AF030699; G3523117; -.			
FT NON_TER 393 393			
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;			

Query Match 100.0%; Score 53; DB 11; Length 393;
Best Local Similarity 100.0%; Pred. No. 7.83e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 267 DQGGESSLS 275

Qy 237 DQGGESSLS 245

RESULT 2
ID Q64906
AC Q64906;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.

OS ALCELAHINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=C500;

RX MEDLINE; 97201573.
RA ENSER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RX MEDLINE; 97404659.
RA ENSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RA ENSER A., PFLANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO ENBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U18243; G1000717; -.
DR EMBL; AF005370; G2337870; -.
SQ SEQUENCE 553 AA; 73645 MW; CCCB53C9 CRC32;

Query Match 100.0%; Score 53; DB 14; Length 653;
Best Local Similarity 100.0%; Pred. No. 7.83e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 291 DOGGESSLS 299
|||||||
QY 237 DOGGESSLS 245

RESULT 3
ID O75326 PRELIMINARY; PRT; 666 AA.
AC O75326;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN L.
GN SEMAL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses.";
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030898; G3523115; -.
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 53; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 7.83e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 269 DOGGESSLS 277
|||||||
QY 237 DOGGESSLS 245

RESULT 4
ID Q05645 PRELIMINARY; PRT; 1032 AA.
AC Q05645;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MAJOR TEGUMENTAL ANTIGEN SM15 PRECURSOR (SM15).
GN GSM70.
OS SCHISTOSOMA MANSONI (BLOOD FLUKE).
OC EUKARYOTA; METAZOA; PLATYHELMINTHES; TREMATODA; DIGenea; STRIGEIDIDA;
OC SCHISTOSOMATOIDEA; SCHISTOSOMATIDAE; SCHISTOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PUERTO RICAN;

RX MEDLINE; 93376023.
RA ABATH F.G., HAGAN P., HOLDER A.A., SIMPSON A.J.;
RT "Structure of the gene encoding a putative Schistosoma mansoni
RT tegumental antigen precursor.";
RL MOL. BIOCHEM. PARASITOL. 60:81-91(1993).
CC -!- FUNCTION: MAY SERVE AS A TARGET OF PROTECTIVE IMMUNITY.
DR EMBL; L07513; G161042; -.
KW ANTIGEN; REPEAT; GLYCOPROTEIN; MYRISTYLATION.
FT REPEAT 460 780 5 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 460 519 R1.
FT REPEAT 520 602 R2.
FT REPEAT 603 684 R3.
FT REPEAT 685 768 R4.
FT REPEAT 769 780 R5.
FT LIPID 43 43 MYRISTATE (POTENTIAL).
FT LIPID 171 171 MYRISTATE (POTENTIAL).
FT LIPID 175 175 MYRISTATE (POTENTIAL).
FT LIPID 177 177 MYRISTATE (POTENTIAL).
FT LIPID 300 300 MYRISTATE (POTENTIAL).
FT CARBOHYD 47 47 POTENTIAL.
FT CARBOHYD 82 82 POTENTIAL.
FT CARBOHYD 85 85 POTENTIAL.
FT CARBOHYD 143 143 POTENTIAL.
FT CARBOHYD 172 172 POTENTIAL.
FT CARBOHYD 179 179 POTENTIAL.
FT CARBOHYD 190 190 POTENTIAL.
FT CARBOHYD 304 304 POTENTIAL.
FT CARBOHYD 354 354 POTENTIAL.
FT CARBOHYD 377 377 POTENTIAL.
FT CARBOHYD 474 474 POTENTIAL.
FT CARBOHYD 557 557 POTENTIAL.
FT CARBOHYD 640 640 POTENTIAL.
FT CARBOHYD 723 723 POTENTIAL.
FT CARBOHYD 816 816 POTENTIAL.
FT CARBOHYD 820 820 POTENTIAL.
FT CARBOHYD 861 861 POTENTIAL.
FT CARBOHYD 881 881 POTENTIAL.
FT CARBOHYD 902 902 POTENTIAL.
FT CARBOHYD 975 975 POTENTIAL.
FT CARBOHYD 986 986 POTENTIAL.
FT CARBOHYD 1001 1001 POTENTIAL.
FT CARBOHYD 1024 1024 POTENTIAL.
SQ SEQUENCE 1032 AA; 116899 MW; F239E4B9 CRC32;

Query Match 83.0%; Score 44; DB 5; Length 1032;
Best Local Similarity 66.7%; Pred. No. 2.30e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 298 EOGGEENLS 306
|||||
QY 237 DOGGESSLS 245

RESULT 5
ID Q93456 PRELIMINARY; PRT; 296 AA.
AC Q93456;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE F10G8.7 PROTEIN.
GN F10G8.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA BASHAM V.;
RL SUBMITTED (SEP-1996) TO ENBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL NATURE 368:32-38(1994).
 DR EMBL; Z80216; E1345455; -.
 SQ SEQUENCE 296 AA: 33613 MW; 5F3E5B16 CRC32;
 Query Match 79.2%; Score 42; DB 5; Length 296;
 Best Local Similarity 66.7%; Pred. No. 7.44e+00;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 214 DQGDSSMS 222
 QY 237 DQGGESSLS 245
 I:::III
 I:::III
 RESULT 6
 ID O53580 PRELIMINARY; PRT; 314 AA.
 AC O53580;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE ORF25 PROTEIN.
 OS LEUCANTA SEPARATA NUCLEAR POLYHEDROSIS VIRUS.
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
 CC NUCLEOPOLYHEDROVIRUS.
 CC [1]
 RN JIN T.;
 RP SEQUENCE FROM N.A.
 RA JIN T.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AB009614; D1025167; -.
 SQ SEQUENCE 314 AA: 36908 MW; 9F35D3B8 CRC32;
 Query Match 79.2%; Score 42; DB 14; Length 314;
 Best Local Similarity 62.5%; Pred. No. 7.44e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 24 EGGNNL 31
 QY 237 DQGGESSL 244
 I:::III
 I:::III
 RESULT 7
 ID Q30228 PRELIMINARY; PRT; 356 AA.
 AC Q30228;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE MHC CLASS I AOTR-G*03 (FRAGMENT).
 GN AOTR-G.
 OS AOTUS TRIVIRGATUS (NIGHT MONKEY) (DOUROUCOULI).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 CC PLATYRRHINI; CEBIDAE; AOTINAE; AOTUS.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 98070787.
 RA CADAVID L.F., SHUFFLEBOTHAM C., RUIZ F.J., YEAGER M., HUGHES A.L.,
 RA WATKINS D.I.;
 RT "Evolutionary instability of the major histocompatibility complex
 RT class I loci in New World primates."
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:14536-14541(1997).
 DR EMBL; U59644; G1389919; -.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PFAM; PF00047; Ig; 1.
 DR PFAM; PF00129; MHC_I; 1.
 KW MHC.
 RN JIN T.;
 RP SEQUENCE FROM N.A.
 RA JIN T.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AB009614; D1025167; -.
 SQ SEQUENCE 314 AA: 36908 MW; 9F35D3B8 CRC32;
 Query Match 79.2%; Score 42; DB 14; Length 314;
 Best Local Similarity 62.5%; Pred. No. 7.44e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 24 EGGNNL 31
 QY 237 DQGGESSL 244
 I:::III
 I:::III
 RESULT 8
 ID Q30226 PRELIMINARY; PRT; 357 AA.
 AC Q30226;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE MHC CLASS I AOTR-G*01 (FRAGMENT).
 GN AOTR-G.
 OS AOTUS TRIVIRGATUS (NIGHT MONKEY) (DOUROUCOULI).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 CC PLATYRRHINI; CEBIDAE; AOTINAE; AOTUS.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 98070787.
 RA CADAVID L.F., SHUFFLEBOTHAM C., RUIZ F.J., YEAGER M., HUGHES A.L.,
 RA WATKINS D.I.;
 RT "Evolutionary instability of the major histocompatibility complex
 RT class I loci in New World primates."
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:14536-14541(1997).
 DR EMBL; U59644; G1389915; -.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PFAM; PF00047; Ig; 1.
 DR PFAM; PF00129; MHC_I; 1.
 KW MHC.
 RN JIN T.;
 RP SEQUENCE FROM N.A.
 RA JIN T.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AB009614; D1025167; -.
 SQ SEQUENCE 357 AA: 39975 MW; BCED1B6 CRC32;
 Query Match 79.2%; Score 42; DB 7; Length 357;
 Best Local Similarity 75.0%; Pred. No. 7.44e+00;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 345 DQGDSSL 352
 QY 237 DQGGESSL 244
 I:::III
 I:::III
 RESULT 9
 ID O45502 PRELIMINARY; PRT; 402 AA.
 AC O45502;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE F39B2.10 PROTEIN.
 GN F39B2.10.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 CC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP DOBSON R.;
 RA DOBSON R.;
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA WATKINS D.I.;
 RT "Evolutionary instability of the major histocompatibility complex
 RT class I loci in New World primates."
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:14536-14541(1997).
 DR EMBL; U59646; G1389919; -.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PFAM; PF00047; Ig; 1.
 DR PFAM; PF00129; MHC_I; 1.
 KW MHC.

FT NON_TER 1
 SQ SEQUENCE 356 AA: 39794 MW; 164A0914 CRC32;
 Query Match 79.2%; Score 42; DB 7; Length 356;
 Best Local Similarity 75.0%; Pred. No. 7.44e+00;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 344 DQGDSSL 351
 QY 237 DQGGESSL 244
 I:::III
 I:::III
 RESULT 8
 ID Q30226 PRELIMINARY; PRT; 357 AA.
 AC Q30226;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE MHC CLASS I AOTR-G*01 (FRAGMENT).
 GN AOTR-G.
 OS AOTUS TRIVIRGATUS (NIGHT MONKEY) (DOUROUCOULI).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 CC PLATYRRHINI; CEBIDAE; AOTINAE; AOTUS.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 98070787.
 RA CADAVID L.F., SHUFFLEBOTHAM C., RUIZ F.J., YEAGER M., HUGHES A.L.,
 RA WATKINS D.I.;
 RT "Evolutionary instability of the major histocompatibility complex
 RT class I loci in New World primates."
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:14536-14541(1997).
 DR EMBL; U59644; G1389915; -.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PFAM; PF00047; Ig; 1.
 DR PFAM; PF00129; MHC_I; 1.
 KW MHC.
 RN JIN T.;
 RP SEQUENCE FROM N.A.
 RA JIN T.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AB009614; D1025167; -.
 SQ SEQUENCE 357 AA: 39975 MW; BCED1B6 CRC32;
 Query Match 79.2%; Score 42; DB 7; Length 357;
 Best Local Similarity 75.0%; Pred. No. 7.44e+00;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 345 DQGDSSL 352
 QY 237 DQGGESSL 244
 I:::III
 I:::III
 RESULT 9
 ID O45502 PRELIMINARY; PRT; 402 AA.
 AC O45502;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE F39B2.10 PROTEIN.
 GN F39B2.10.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 CC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP DOBSON R.;
 RA DOBSON R.;
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA WATKINS D.I.;
 RT "Evolutionary instability of the major histocompatibility complex
 RT class I loci in New World primates."
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:14536-14541(1997).
 DR EMBL; U59646; G1389919; -.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PFAM; PF00047; Ig; 1.
 DR PFAM; PF00129; MHC_I; 1.
 KW MHC.

```
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC
CC DNAJ-LIKE PROTEINS.
DR EMBL: Z92834; E1346583; -.
DR PROSITE: PS00636; DNAJ.1; 1.
DR PROSITE: PS00637; DNAJ_CXXCXGXG; 1.
RW CHAPERONE: DNA REPLICATION.
SQ SEQUENCE 402 AA; 44308 MW; 592FB9CF CRC32;

Query Match 79.2%; Score 42; DB 5; Length 402;
Best Local Similarity 75.0%; Pred. No. 7.44e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 65 DQGEEAL 72
QY 237 DQGESL 244

RESULT 10
ID Q98787 PRELIMINARY; PRT; 150 AA.
AC Q98787;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOPROTEIN (FRAGMENT).
OS VESICULAR STOMATITIS VIRUS.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES; RHABDOVIRIDAE;
OC VESICULOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1187CRB2;
RX MEDLINE: 97075113.
RA RODRIGUEZ L.L., FITCH W.M., NICHOL S.T.;
RT "Ecological factors rather than temporal factors dominate the
RT evolution of vesicular stomatitis virus.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:13030-13035(1996).
DR EMBL: U39219; G1628519; -.
DR PFAM: PF00922; Phosphoprotein; 1.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16625 MW; 88E7B19D CRC32;

Query Match 75.5%; Score 40; DB 14; Length 150;
Best Local Similarity 66.7%; Pred. No. 2.31e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 23 ENGERSLS 31
QY 237 DQGESLS 245

RESULT 11
ID Q98781 PRELIMINARY; PRT; 150 AA.
AC Q98781;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOPROTEIN (FRAGMENT).
OS VESICULAR STOMATITIS VIRUS.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES; RHABDOVIRIDAE;
OC VESICULOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-0988CRB;
RX MEDLINE: 97075113.
RA RODRIGUEZ L.L., FITCH W.M., NICHOL S.T.;
RT "Ecological factors rather than temporal factors dominate the
RT evolution of vesicular stomatitis virus.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:13030-13035(1996).
DR EMBL: U39220; G1628521; -.
DR PFAM: PF00922; Phosphoprotein; 1.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16785 MW; A89BC9AB CRC32;
```

```
RL PROC. NATL. ACAD. SCI. U.S.A. 93:13030-13035(1996).
DR EMBL: U39213; G1628507; -.
DR PFAM: PF00922; Phosphoprotein; 1.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16785 MW; A89BC9AB CRC32;

Query Match 75.5%; Score 40; DB 14; Length 150;
Best Local Similarity 66.7%; Pred. No. 2.31e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 23 ENGERSLS 31
QY 237 DQGESLS 245

RESULT 12
ID Q98783 PRELIMINARY; PRT; 150 AA.
AC Q98783;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOPROTEIN (FRAGMENT).
OS VESICULAR STOMATITIS VIRUS.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES; RHABDOVIRIDAE;
OC VESICULOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-0988CRB8;
RX MEDLINE: 97075113.
RA RODRIGUEZ L.L., FITCH W.M., NICHOL S.T.;
RT "Ecological factors rather than temporal factors dominate the
RT evolution of vesicular stomatitis virus.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:13030-13035(1996).
DR EMBL: U39215; G1628511; -.
DR PFAM: PF00922; Phosphoprotein; 1.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16785 MW; A89BC9AB CRC32;

Query Match 75.5%; Score 40; DB 14; Length 150;
Best Local Similarity 66.7%; Pred. No. 2.31e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 23 ENGERSLS 31
QY 237 DQGESLS 245

RESULT 13
ID Q98788 PRELIMINARY; PRT; 150 AA.
AC Q98788;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOPROTEIN (FRAGMENT).
OS VESICULAR STOMATITIS VIRUS.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES; RHABDOVIRIDAE;
OC VESICULOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1187CRB3;
RX MEDLINE: 97075113.
RA RODRIGUEZ L.L., FITCH W.M., NICHOL S.T.;
RT "Ecological factors rather than temporal factors dominate the
RT evolution of vesicular stomatitis virus.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:13030-13035(1996).
DR EMBL: U39220; G1628521; -.
DR PFAM: PF00922; Phosphoprotein; 1.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16785 MW; A89BC9AB CRC32;
```

Query Match 75.5%; Score 40; DB 14; Length 150;
Best Local Similarity 66.7%; Pred. No. 2.31e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Search completed: Thu Jul 8 18:40:11 1999
Job time : 12 secs.

Db 23 ENGGEKSL 31
QY 237 DOGESLS 245

RESULT 14
ID Q98782 PRELIMINARY; PRT; 150 AA.
AC Q98782;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOPROTEIN (FRAGMENT).
OS VESICULAR STOMATITIS VIRUS.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES; RHABDOVIRIDAE;
OC VESICULOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-0988CRB2;
RX MEDLINE; 97075113.
RA RODRIGUEZ L.L., FITCH W.M., NICHOL S.T.;
RT "Ecological factors rather than temporal factors dominate the
evolution of vesicular stomatitis virus";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:13030-13035(1996).
DR EMBL; U39214; G1628509; -
DR PFW; PF00922; Phosphoprotein; 1.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16785 MW; A89BC9AB CRC32;

Query Match 75.5%; Score 40; DB 14; Length 150;
Best Local Similarity 66.7%; Pred. No. 2.31e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 23 ENGGEKSL 31
QY 237 DOGESLS 245

RESULT 15
ID Q98785 PRELIMINARY; PRT; 150 AA.
AC Q98785;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOPROTEIN (FRAGMENT).
OS VESICULAR STOMATITIS VIRUS.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES; RHABDOVIRIDAE;
OC VESICULOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1080CRB;
RX MEDLINE; 97075113.
RA RODRIGUEZ L.L., FITCH W.M., NICHOL S.T.;
RT "Ecological factors rather than temporal factors dominate the
evolution of vesicular stomatitis virus";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:13030-13035(1996).
DR EMBL; U39217; G1628515; -
DR PFW; PF00922; Phosphoprotein; 1.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16785 MW; A89BC9AB CRC32;

Query Match 75.5%; Score 40; DB 14; Length 150;
Best Local Similarity 66.7%; Pred. No. 2.31e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 23 ENGGEKSL 31
QY 237 DOGESLS 245

THIS PAGE BLANK (USPTO)

WIREH (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:58:42 1999; Maspar time 10.35 Seconds
Tabular output not generated. 16,436 Million cell updates/sec

Title: >US-09-041-236-2
Description: (475-482) from US09041236.pep (25 of 45)
Perfect Score: 79
Sequence: 1 RDPYCGWD 8

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266508 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 18.022; Variance 56.032; scale 0.322

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	79	100.0	724	13	Drosophila semaphorin	1.24e-01
2	73	92.4	477	13	Human collapsin	5.94e-01
3	73	92.4	771	13	Human semaphorin III	5.94e-01
4	73	92.4	775	33	Human semaphorin	5.94e-01
5	73	92.4	974	33	Human secreted protei	5.94e-01
6	67	84.8	861	22	Mouse CD100 antigen	2.77e+00
7	67	84.8	861	32	Human semaphorin	2.77e+00
8	67	84.8	862	22	Human CD100 antigen	2.77e+00
9	65	82.3	650	13	Drosophila semaphorin	4.59e+00
10	65	82.3	730	13	Grasshopper semaphori	4.59e+00
11	65	82.3	730	13	Rat semaphorin z	4.59e+00
12	65	82.3	888	25	Human semaphorin z	4.59e+00
13	65	82.3	930	32	Human semaphorin Y	4.59e+00
14	64	81.0	929	32	Rat semaphorin Y	5.90e+00
15	56	70.9	118	10	Epitope of Blastomyce	4.23e+01
16	56	70.9	313	1	P28II antigen	4.23e+01

17	53	67.1	712	13	R71384	Tribolium semaphorin	8.68e-01
18	51	64.6	7	13	R70412	Semaphorin binding pe	1.39e-02
19	51	64.6	10	13	R70415	Semaphorin binding pe	1.39e-02
20	51	64.6	11	13	R70414	Semaphorin binding pe	1.39e-02
21	51	64.6	12	13	R70413	Semaphorin binding pe	1.39e-02
22	51	64.6	35	30	W53659	Tarantula venom produ	1.39e-02
23	51	64.6	39	30	W53658	Tarantula venom produ	1.39e-02
24	51	64.6	42	30	W53662	Tarantula venom produ	1.39e-02
25	51	64.6	88	30	W50153	Tarantula venom produ	1.39e-02
26	51	64.6	198	19	W04252	Polytropic MX27 provi	1.39e-02
27	51	64.6	209	19	W04250	10A1 murine leukaemia	1.39e-02
28	51	64.6	376	24	W32863	Saccharopolyspora mod	1.39e-02
29	50	63.3	39	30	W53657	Tarantula venom produ	1.76e-02
30	50	63.3	41	30	W53660	Tarantula venom produ	1.76e-02
31	50	63.3	42	30	W53661	Tarantula venom produ	1.76e-02
32	50	63.3	42	30	W53663	Tarantula venom produ	1.76e-02
33	50	63.3	48	1	P93361	Amino acid sequence e	1.76e-02
34	50	63.3	83	30	W50156	Tarantula venom produ	1.76e-02
35	50	63.3	87	30	W50154	Tarantula venom produ	1.76e-02
36	50	63.3	87	30	W53654	Tarantula venom produ	1.76e-02
37	50	63.3	87	30	W53656	Tarantula venom produ	1.76e-02
38	50	63.3	122	8	R39346	Phospholipase A2 subu	1.76e-02
39	50	63.3	137	12	R63045	RPLA2-10	1.76e-02
40	50	63.3	138	12	R63046	HPLA2-10	1.76e-02
41	50	63.3	360	34	W69597	Mouse G-protein coupl	1.76e-02
42	50	63.3	512	3	P50630	Human pancreatic alph	1.76e-02
43	50	63.3	512	3	P50629	Human parotid alpha-a	1.76e-02
44	49	62.0	387	37	W81589	Protein encoded by hu	2.21e+02
45	48	60.8	201	32	W53966	Volutella collectotri	2.79e+02

ALIGNMENTS

RESULT 1
ID R71383 standard; Protein: 724 AA.
AC R71383;
DT 21-NOV-1995 (first entry)
DE Drosophila semaphorin II protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; human; vaccinia receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Drosophila sp.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; Q87445.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 79-82; 101pp; English.
CC The sequence of the Drosophila semaphorin II protein. The gene was
CC isolated by PCR using primers based on sequence homology between the
CC complete grasshopper and partial Tribolium and Manduca (moth) semaphorin
CC gene sequences. The products gave two different sequences, each of which
CC was used to obtain its respective complete sequence: semaphorin I
CC (Q87444) and II. The proteins encoded by the grasshopper semaphorin I
CC (Q87441), human semaphorin III (Q87442), vaccinia virus semaphorin IV,
CC Drosophila semaphorin I and II, Tribolium semaphorin I (Q87446) or
CC variola major (smallpox) virus semaphorin IV (Q87447) genes were used to
CC generate a series of peptides (R70370-R70418), which retain semaphorin
CC receptor binding activity. The semaphorin derived or semaphorin receptor
CC derived peptides are potent modulators of nerve cell growth, immune
CC responsiveness and viral pathogenesis. They can be used in diagnosis and
CC treatment of neurological disease and neuro-regeneration, immune
CC modulation and diagnosis and treatment of viral and oncological infection
CC and diseases.
SQ Sequence 724 AA;

```

Query Match      100.0%; Score 79; DB 13; Length 724;
Best Local Similarity 100.0%; Pred. No. 1.24e-01;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Db      537 rdpvcgwd 544
Qy      475 RDPYCGWD 482

RESULT      2
ID      R74175 standard; Protein; 477 AA.
AC      R74175;
DE      01-NOV-1995 (first entry)
DE      Human collapsin.
KW      Collapsin; antibody; therapy.
OS      Homo sapiens.
FH      Key      Location/Qualifiers
FT      binding_site      9..19 "antibody binding site"
FT      binding_site      51..65
FT      /note= "antibody binding site"
PN      US5416197-A.
PD      16-MAY-1995.
PF      15-OCT-1993; 136922.
PR      15-OCT-1993; US-136922.
PA      (UYPE-) UNIV PENNSYLVANIA.
PI      Luo Y, Raper JA;
DR      WPI: 95-193478/25.
DR      N-PSDB: Q22331.
PT      New antibody to human collapsin - used to inhibit the activity of
PT      collapsin, to induce neurite out-growth and to treat individuals with
PT      nerve damage.
PS      Claim 2; Columns 15-18; 11pp; English.
CC      An antibody capable of specifically binding at least a portion of
CC      the collapsin protein can be used to purify human collapsin and
CC      to inhibit the activity of the protein. It can be used to induce
CC      neurite outgrowth by neuronal cells and to treat individuals
CC      suffering from nerve damage.
SQ      Sequence      477 AA;

Query Match      92.4%; Score 73; DB 13; Length 477;
Best Local Similarity 87.5%; Pred. No. 5.94e-01;
Matches      7; Conservative      1; Mismatches      0; Indels      0; Gaps      0;

Db      299 rdpvcawd 306
Qy      475 RDPYCGWD 482

RESULT      3
ID      R71380 standard; Protein; 771 AA.
AC      R71380;
DE      21-NOV-1995 (first entry)
DE      Human semaphorin III protein.
KW      Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW      variola major virus; smallpox; semaphorin receptor binding activity;
KW      modulation; nerve cell growth; immune response; viral pathogenesis;
KW      neurological disease; neuro-regeneration; oncological infection.
OS      Homo sapiens.
PN      WO9507706-A.
PD      23-MAR-1995.
PF      13-SEP-1994; U10151.
PR      13-SEP-1993; US-121713.
PA      (REGC ) UNIV CALIFORNIA.
PI      Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI      O'Connor T;
DR      WPI: 95-131177/17.
DR      N-PSDB: Q87442.
PT      New class of semaphorin peptide(s) and polypeptide(s) - are
PT      potent modulators of nerve cell growth and regeneration
PS      Example 2; Page 60-63; 101pp; English.
CC      The sequence of the human semaphorin III protein. The proteins
CC      encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,

```

```

CC      vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC      (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC      virus semaphorin IV (Q87447) genes were used to generate a series of
CC      peptides (R70370-R70418), which retain semaphorin receptor binding
CC      activity. The semaphorin derived or semaphorin receptor derived peptides
CC      are potent modulators of nerve cell growth, immune responsiveness and
CC      viral pathogenesis. They can be used in diagnosis and treatment of
CC      neurological disease and neuro-regeneration, immune modulation and
CC      diagnosis and treatment of viral and oncological infection and diseases.
SQ      Sequence      771 AA;

Query Match      92.4%; Score 73; DB 13; Length 771;
Best Local Similarity 87.5%; Pred. No. 5.94e-01;
Matches      7; Conservative      1; Mismatches      0; Indels      0; Gaps      0;

Db      531 rdpvcawd 538
Qy      475 RDPYCGWD 482

RESULT      4
ID      W63748 standard; Protein; 775 AA.
AC      W63748;
DE      01-OCT-1998 (first entry)
DE      Human semaphorin.
KW      Human; semaphorin; diagnosis; nerve disease; nerve growth inhibitor;
KW      neurological disease; atopic skin inflammation; autoimmune disease;
KW      pain.
OS      Homo sapiens.
PN      WO9822504-A1.
PD      28-MAY-1998.
PF      12-NOV-1997; J04111.
PR      15-NOV-1996; JP-321068.
PA      (SUMU ) SUMITOMO PHARM CO LTD.
PI      Furuyama T, Inagaki S;
DR      WPI: 98-312416/27.
DR      N-PSDB: V35367.
PT      Gene encoding new semaphorin nerve growth inhibitor - useful in
PT      diagnosis, treatment and study of neurological diseases
PS      Claim 1; Page 33-37; 49pp; Japanese.
CC      The present sequence represents human semaphorin, a nerve growth
CC      inhibitor. The semaphorin protein, and gene encoding the protein,
CC      and their derivatives, are used in the diagnosis, treatment and
CC      study of neurological disorders such as atopic skin inflammation,
CC      autoimmune diseases and pain.
SQ      Sequence      775 AA;

Query Match      92.4%; Score 73; DB 33; Length 775;
Best Local Similarity 87.5%; Pred. No. 5.94e-01;
Matches      7; Conservative      1; Mismatches      0; Indels      0; Gaps      0;

Db      533 rdpvcawd 540
Qy      475 RDPYCGWD 482

RESULT      5
ID      W64221 standard; Protein; 974 AA.
AC      W64221;
DE      06-OCT-1998 (first entry)
DE      Human secreted protein from clone CUI45_1.
KW      Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW      cell proliferation; differentiation; immune system; suppressor; ligand;
KW      regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW      chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW      anti-inflammatory.
OS      Homo sapiens.
PN      WO9827205-A2.
PD      25-JUN-1998.
PF      17-DEC-1997; U23330.
PR      16-DEC-1997; US-991872.
PR      18-DEC-1996; US-769192.
PR      13-JAN-1997; US-783401.

```

```

PA (GEM ) GENETICS INST INC.
PI Acostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-362774/31.
DR N-PSDB; V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT fetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CUL45_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 92.4%; Score 73; DB 33; Length 974;
Best Local Similarity 100.0%; Pred. No. 5.94e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 528 rdpvcgw 534
QY 475 RDPYCGW 481

RESULT 6
ID W17658 standard; Protein; 861 AA.
AC W17658;
DT 24-JUL-1997 (first entry)
DE Mouse CD100 antigen.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
OS Mus sp.
FH Key Location/Qualifiers
FT peptide 1..41
FT /label= Sig_peptide
FT protein 42..861
FT /label= Mat_protein
FT domain 42..553
FT /label= Semaphorin_domain
FT domain 554..630
FT /label= Ig-like_domain
FT domain 631..732
FT /label= Stalk_domain
FT domain 734..752
FT /label= Transmembrane_domain
FT domain 753..861
FT /label= Cytoplasmic_domain
FT modified_site 807..814
FT /label= Phosphorylation
FT /note= "putative tyrosine phosphorylation site"
PN W09717368-A1.
PD 15-MAY-1997.
PF 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND ) DANA FARBER CANCER INST.
PI Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB; T60666.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Example 8; Page 86-89; 13pp; English.
CC Mouse CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60666) isolated from murine T cells. Human CD100 antigen
CC (W17657) has also been identified. CD100 polypeptides and fusion

```

```

CC proteins, nucleic acids, and host cells expressing CD100 can be
CC utilised in diagnostic and therapeutic methods involving modulation
CC of B and T cell responses, neuron axonal growth and immune cell-
CC nerve cell interaction.
SQ Sequence 861 AA;

Query Match 84.8%; Score 67; DB 22; Length 861;
Best Local Similarity 85.7%; Pred. No. 2.77e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 516 rdpvcaw 522
QY 475 RDPYCGW 481

RESULT 7
ID W58540 standard; Protein; 861 AA.
AC W58540;
DT 02-SEP-1998 (first entry)
DE Human semaphorin.
KW Human; semaphorin; diagnosis; nervous disease; immune disease;
KW nerve extension inhibiting activity.
OS Homo sapiens.
PN J10155490-A.
PD 16-JUN-1998.
PF 27-NOV-1996; 332900.
PR 27-NOV-1996; JP-332900.
PA (SUMU ) SUMITOMO SEIYAKU KK.
DR WPI: 98-391044/34.
DR N-PSDB; V31121.
PT New human semafolin gene - useful in the diagnosis of nervous system
PT and immune disorders
PS Claim 1; Page 10-12; 15pp; Japanese.
CC The present sequence represents human semaphorin (translated from the
CC Japanese specification as semafolin). Semaphorin has nerve extension
CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
CC treatment and researches on nervous diseases and immune diseases.
SQ Sequence 861 AA;

Query Match 84.8%; Score 67; DB 32; Length 861;
Best Local Similarity 85.7%; Pred. No. 2.77e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 516 rdpvcaw 522
QY 475 RDPYCGW 481

RESULT 8
ID W17657 standard; Protein; 862 AA.
AC W17657;
DT 24-JUL-1997 (first entry)
DE Human CD100 antigen.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
KW vaccine.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..41
FT /label= Sig_peptide
FT protein 42..862
FT /label= Mat_protein
FT domain 42..553
FT /label= Semaphorin_domain
FT domain 554..630
FT /label= Ig-like_domain
FT domain 631..733
FT /label= Stalk_domain
FT domain 735..752
FT /label= Transmembrane_domain
FT domain 753..862
FT /label= Cytoplasmic_domain
FT modified_site 808..815
FT /label= Phosphorylation

```

FT W09717368-A1. /note= "putative tyrosine phosphorylation site"

PN 15-MAY-1997.
PD 12-NOV-1996: U18645.
PR 09-NOV-1995: US-556422.
PA (DAND) DANA FARBER CANCER INST.
PI Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB: T60665.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Claim 7: Page 70-72; 135pp; English.
CC Human CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60665) isolated by COS cell expression cloning. CD100
CC polypeptides and fusion proteins, nucleic acids, and host cells
CC expressing CD100 can be utilised in diagnostic and therapeutic
CC methods involving modulation of B and T cell responses, neuron
CC axonal growth and immune cell-nerve cell interaction.
SQ Sequence 862 AA;

Query Match 84.8%; Score 67; DB 22; Length 862;
Best Local Similarity 85.7%; Pred. No. 2.77e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 516 rdpycaw 522
QY 475 RDPYCGW 481

RESULT 9

ID R71382 standard; Protein: 650 AA.

AC R71382; 1995 (first entry)
DE Drosophila semaphorin I protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Drosophila sp.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB: Q87444.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2: Page 74-77; 101pp; English.
CC The sequence of the Drosophila semaphorin I protein. The gene was
CC isolated by PCR using primers based on sequence homology between the
CC complete grasshopper and partial Tribolium and Manduca (moth) semaphorin
CC gene sequences. The products gave two different sequences, each of which
CC was used to obtain its respective complete sequence: semaphorin I and II
CC (Q87445). The proteins encoded by the grasshopper semaphorin I (Q87441),
CC human semaphorin III (Q87442), vaccinia virus semaphorin IV, Drosophila
CC semaphorin I and II, Tribolium semaphorin I (Q87446) or variola major
CC (smallpox) virus semaphorin IV (Q87447) genes were used to generate a
CC series of peptides (R70370-R70418), which retain semaphorin receptor
CC binding activity. The semaphorin derived or semaphorin receptor derived
CC peptides are potent modulators of nerve cell growth, immune
CC responsiveness and viral pathogenesis. They can be used in diagnosis and
CC treatment of neurological disease and neuro-regeneration, immune
CC modulation and diagnosis and treatment of viral and oncological infection
CC and diseases.
SQ Sequence 650 AA;

Query Match 82.3%; Score 65; DB 13; Length 650;
Best Local Similarity 75.0%; Pred. No. 4.59e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 414 qdpycawd 421
QY 475 RDPYCGWD 482

RESULT 10

ID R71379 standard; Protein: 730 AA.
AC R71379;
DT 21-NOV-1995 (first entry)
DE Grasshopper semaphorin I protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Grasshopper sp.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB: Q87441.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 1: Page 68-72; 101pp; English.
CC The sequence of the grasshopper semaphorin I protein. The proteins
CC encoded by the grasshopper semaphorin I, human semaphorin III (Q87442),
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 730 AA;

Query Match 82.3%; Score 65; DB 13; Length 730;
Best Local Similarity 75.0%; Pred. No. 4.59e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 508 qdpycawd 515
QY 475 RDPYCGWD 482

RESULT 11

ID W19856 standard; Protein: 887 AA.
AC W19856;
DT 13-FEB-1998 (first entry)
DE Rat semaphorin 2.
KW Semaphorin 2; central nerve extension; rat; human; inhibitor;
KW central nerve regeneration promoter.
OS Rattus norvegicus.
PN W09720928-A1.
PD 12-JUN-1997.
PF 02-DEC-1996; J03517.
PR 31-OCT-1996; JP-307205.
PR 06-DEC-1995; JP-345187.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 97-319775/29.
DR N-PSDB: T72107.
PT Semaphorin 2 and the gene encoding it - also inhibitors of its
PT action which can be used as promoters of central nerve regeneration
PS Claim 3; Page 66-71; 106pp; Japanese.

CC This sequence represents the rat semaphorin 2. This protein sequence, and
CC the human semaphorin 2 protein (see W19857) can be used for screening
CC possible candidates for activity as semaphorin 2 inhibitors. Proteins
CC containing all or part of the semaphorin 2 sequence are useful as central
CC nerve extension inhibitors. Semaphorin 2 inhibitors identified by
CC screening can be used as promoters of central nerve regeneration.
SQ Sequence 887 AA;

Query Match 82.3%; Score 65; DB 25; Length 887;
Best Local Similarity 85.7%; Pred. No. 4.59e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 542 qdpycgw 548
:|||||
QY 475 RDPYCGW 481

RESULT 12

ID W19857 standard; Protein; 888 AA.

AC W19857;

DT 13-FEB-1998 (first entry)

DE Human semaphorin 2.

KW Semaphorin 2; central nerve extension; rat; human; inhibitor;

OS Central nerve regeneration promoter.

OS Homo sapiens.

PN W09720928-A1.

PD 12-JUN-1997.

PF 02-DEC-1996; J03517.

PR 31-OCT-1996; JP-307205.

PR 06-DEC-1995; JP-345187.

PA (SUMU) SUMITOMO PHARM CO LTD.

PI Kikuchi K, Kimura T;

DR WPI: 97-319775/29.

DR N-PSDB; T72108.

PT Semaphorin 2 and the gene encoding it - also inhibitors of its

PT action which can be used as promoters of central nerve regeneration

PS Claim 3; Page 77-82; 105pp; Japanese.

CC This sequence represents the human semaphorin 2. This protein sequence,

CC and the rat semaphorin 2 protein (see W19856) can be used for screening

CC possible candidates for activity as semaphorin 2 inhibitors. Proteins

CC containing all or part of the semaphorin 2 sequence are useful as central

CC nerve extension inhibitors. Semaphorin 2 inhibitors identified by

CC screening can be used as promoters of central nerve regeneration.

SQ Sequence 888 AA;

Query Match 82.3%; Score 65; DB 25; Length 888;
Best Local Similarity 85.7%; Pred. No. 4.59e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 540 qdpycgw 546
:|||||
QY 475 RDPYCGW 481

RESULT 13

ID W57260 standard; Protein; 930 AA.

AC W57260;

DT 02-SEP-1998 (first entry)

DE Human semaphorin Y.

KW Human; semaphorin Y; nerve extension inhibitor; central nervous system;

KW peripheral nerve growth.

OS Homo sapiens.

PN W09811216-A1.

PD 19-MAR-1998.

PF 09-SEP-1997; J03167.

PR 08-AUG-1997; JP-227220.

PR 11-SEP-1996; JP-263565.

PA (SUMU) SUMITOMO PHARM CO LTD.

PI Kikuchi K, Kimura T;

DR WPI: 98-250958/22.

DR N-PSDB; V28915, V28916.

PT DNA encoding human and rat semaphorin Y - an inhibitor of nerve

PT extension

PS Claim 1; Page 65-70; 85pp; Japanese.
CC The present sequence represents human semaphorin Y, which inhibits
CC nerve extension. Semaphorin Y genes and proteins may be used to inhibit
CC peripheral nerve growth. Semaphorin Y antagonists can be used to
CC accelerate regrowth of the central nervous system.
SQ Sequence 930 AA;

Query Match 82.3%; Score 65; DB 32; Length 930;
Best Local Similarity 85.7%; Pred. No. 4.59e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 533 qdpycgw 539
:|||||
QY 475 RDPYCGW 481

RESULT 14

ID W57259 standard; Protein; 929 AA.

AC W57259;

DT 02-SEP-1998 (first entry)

DE Rat semaphorin Y.

KW Rat; semaphorin Y; nerve extension inhibitor; central nervous system;

KW peripheral nerve growth.

OS Rattus norvegicus.

PN W09811216-A1.

PD 19-MAR-1998.

PF 09-SEP-1997; J03167.

PR 08-AUG-1997; JP-227220.

PR 11-SEP-1996; JP-263565.

PA (SUMU) SUMITOMO PHARM CO LTD.

PI Kikuchi K, Kimura T;

DR WPI: 98-250958/22.

DR N-PSDB; V28913, V28914.

PT DNA encoding human and rat semaphorin Y - an inhibitor of nerve

PT extension

PS Claim 1; Page 54-58; 85pp; Japanese.

CC The present sequence represents rat semaphorin Y, which inhibits nerve

CC extension. Semaphorin Y genes and proteins may be used to inhibit

CC peripheral nerve growth. Semaphorin Y antagonists can be used to

CC accelerate regrowth of the central nervous system.

SQ Sequence 929 AA;

Query Match 81.0%; Score 64; DB 32; Length 929;
Best Local Similarity 100.0%; Pred. No. 5.90e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 534 dpycgw 539
:|||||
QY 476 DPYCGW 481

RESULT 15

ID R51354 standard; Protein; 118 AA.

AC R51354;

DT 27-OCT-1994 (first entry)

DE Epitope of Blastomyces dermatitidis WI-1 protein.

KW Fungus; infection; vaccine; treatment; diagnosis; detection; assay;

KW Blastomyces dermatitidis.

OS Blastomyces dermatitidis.

PN US5302530-A.

PD 12-APR-1994.

PF 16-FEB-1993; 021537.

PR 16-FEB-1993; US-021537.

PA (WISC) WISCONSIN ALUMNI RES ASSOC.

PI Hogan LH, Jones JM, Klein BS;

DR WPI: 94-117608/14.

DR N-PSDB; Q45434.

PT Recombinant vector contg. cDNA encoding an epitope of Blastomyces

PT dermatitidis WI-1 protein - used in assays for this fungus and in

PT vaccines

PS Disclosure; Column 11-12; 9pp; English.

CC The DNA encoding this polypeptide can be used in an improved assay

CC for detecting the presence of Blastomyces dermatitidis. This

CC antigenic polypeptide can be used in diagnostic tests for
CC B. dermatitidis and in vaccines against the fungus.
SQ Sequence 118 AA;

Query Match 70.9%; Score 56; DB 10; Length 118;
Best Local Similarity 66.7%; Pred. No. 4.23e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Db 18 kdpyncdwd 26

QY 475 RDPY-CGWD 482
:||| |||

Search completed: Thu Jul 8 18:58:57 1999
Job time : 15 secs.

WQREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:57:41 1999; MasPar time 7.74 Seconds
Tabular output not generated. 56.378 Million cell updates/sec

Title: >US-09-041-236-2
Description: (475-482) from US09041236.pep (25 of 45)
Perfect Score: 79
Sequence: 1 RDPYCGWD 8

Scoring table:
PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 24.263; Variance 32.137; scale 0.755

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	79	100.0	666	4	SEMAPHORIN L.	1.98e-05
2	79	100.0	706	5	SEMAPHORIN-II.	1.98e-05
3	79	100.0	1093	11	SEMAPHORIN G PRECURSOR	1.98e-05
4	73	92.4	748	11	SEMAPHORIN A PRECURSOR	4.83e-04
5	73	92.4	749	4	SEMAPHORIN V.	4.83e-04
6	73	92.4	750	4	SEMAPHORIN V.	4.83e-04
7	73	92.4	751	13	COLLAPSPIN 3.	4.83e-04
8	73	92.4	751	4	SEMAPHORIN E.	4.83e-04
9	73	92.4	751	11	SEMAPHORIN E. PRECURSOR	4.83e-04
10	73	92.4	753	4	SEMAPHORIN III FAMILY	4.83e-04
11	73	92.4	754	11	SEMAPHORIN IV ISOFORM	4.83e-04
12	73	92.4	761	13	COLLAPSPIN-2.	4.83e-04
13	73	92.4	771	4	SEMAPHORIN-III.	4.83e-04
14	73	92.4	772	11	SEMAPHORIN D PRECURSOR	4.83e-04
15	73	92.4	772	13	COLLAPSPIN.	4.83e-04
16	73	92.4	772	11	SEMAPHORIN III/COLLAPS	4.83e-04
17	73	92.4	775	4	KIAA0331.	4.83e-04
18	73	92.4	775	11	SEMAPHORIN H.	4.83e-04
19	73	92.4	785	13	COLLAPSPIN 5.	4.83e-04
20	73	92.4	785	11	SEMAPHORIN IV ISOFORM	4.83e-04

21	73	92.4	785	4	Q15704	SEMAPHORIN.	4.83e-04
22	73	92.4	785	4	Q13275	SEMAPHORIN IV.	4.83e-04
23	73	92.4	888	11	O35464	SEMAPHORIN VIA.	4.83e-04
24	73	92.4	2135	4	O43157	KIAA0407.	4.83e-04
25	71	89.9	1074	4	Q13591	SEMAPHORIN F HOMOLOG.	1.37e-03
26	71	89.9	1077	11	Q62217	SEMAPHORIN F PRECURSOR	1.37e-03
27	70	88.6	834	11	O64151	SEMAPHORIN I (N-SEMA F	2.29e-03
28	67	84.8	295	13	Q90665	COLLAPSPIN-4 (FRAGMENT)	1.05e-02
29	67	84.8	299	13	Q90666	COLLAPSPIN-5 (FRAGMENT)	1.05e-02
30	67	84.8	782	11	Q62179	SEMAPHORIN C (SEM C)	1.05e-02
31	67	84.8	861	11	Q09126	SEMAPHORIN J (SEMAPHOR	1.05e-02
32	67	84.8	862	4	Q92854	SEMAPHORIN.	1.05e-02
33	65	82.3	446	4	O60650	HYPOPHETICAL 45.7 KD P	2.87e-02
34	65	82.3	730	5	Q26473	FASCICLIN IV.	2.87e-02
35	65	82.3	771	5	Q24322	SEMAPHORIN-I.	2.87e-02
36	65	82.3	886	11	O54951	SEMAPHORIN N (SEMAPHOR	2.87e-02
37	65	82.3	887	11	O70141	SEMAPHORIN Z.	2.87e-02
38	65	82.3	1841	4	O15031	KIAA0315 PROTEIN (FRAG	2.87e-02
39	62	78.5	755	2	O68483	REGULATORY PROTEIN NOS	1.25e-01
40	61	77.2	401	5	O17186	F34D6.4 PROTEIN.	2.02e-01
41	61	77.2	760	11	Q62178	SEMAPHORIN B PRECURSOR	2.02e-01
42	61	77.2	1894	11	P70206	PLEXIN 1.	2.02e-01
43	61	77.2	1905	13	Q91823	REGULATORY PROTEIN NOS	3.26e-01
44	60	75.9	731	2	O68476	XANTHOMONAS COMPESTRIS	8.36e-01
45	58	73.4	617	2	Q56794		

ALIGNMENTS

RESULT 1
ID O75326 PRELIMINARY; PRT; 666 AA.
AC O75326;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN L.
GN SEMAL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030698; G3523115;
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 79; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 1.98e-05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 507 RDPYCGWD 514
Qy 475 RDPYCGWD 482

RESULT 2
ID Q24323 PRELIMINARY; PRT; 706 AA.
AC Q24323;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN-II.
GN DSEMA-II.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.

```
RA KOLODKIN A.L., MATTHES D.J., GOODMAN C.S.;
RL CELL 0:0-0(0);
DR EMBL; L26083; G436557; -.
DR FLYBASE; FBgn0011260; sema-II.
SQ SEQUENCE 706 AA; 81023 MW; B4DB8627 CRC32;

Query Match 100.0%; Score 79; DB 5; Length 706;
Best Local Similarity 100.0%; Pred. No. 1.98e-05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 519 RDPYCGWD 526
QY 475 RDPYCGWD 482

RESULT 3
ID Q60519 PRELIMINARY; PRT; 1093 AA.
AC Q60519;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN G PRECURSOR (SEMAPHORIN G).
GN SEMAG OR SEMG.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI;
RX MEDLINE; 96414430.
RA ADAM R.H., BETZ H., PUSCHEL A.W.;
RT "A novel class of murine semaphorins with homology to thrombospondin
is differentially expressed during early embryogenesis.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI;
RX MEDLINE; 96414430.
RA ADAM R.H.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X97818; E244976; -.
DR MGD; MGI:107555; SEMAG.
DR PFAM; PF00090; tSp_1; 5.
KW SIGNAL.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1093 POTENTIAL.
SQ SEQUENCE 1093 AA; 120325 MW; F009708B CRC32;

Query Match 100.0%; Score 79; DB 11; Length 1093;
Best Local Similarity 100.0%; Pred. No. 1.98e-05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 511 RDPYCGWD 518
QY 475 RDPYCGWD 482

RESULT 4
ID Q62177 PRELIMINARY; PRT; 748 AA.
AC Q62177;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN A PRECURSOR (SEM A).
GN SEMAA OR SEMA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI;
RX MEDLINE; 95267431.
RA PUSCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine sema, a murin D/collapsin is a member of a diverse gene family
and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRIO.
CC LOW LEVELS FOUND BETWEEN DAYS 10-12.
CC EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL
BIRTH.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85990; G854324; -.
DR MGD; MGI:107561; SEMAA.
DR PFAM; PF00047; Ig; 1.
KW SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 748 SEMAPHORIN A.
FT DOMAIN 586 649 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 748 AA; 82894 MW; A7E53A8D CRC32;

Query Match 92.4%; Score 73; DB 11; Length 748;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 529 RDPYCAWD 536
QY 475 RDPYCGWD 482

RESULT 5
ID Q13214 PRELIMINARY; PRT; 749 AA.
AC Q13214;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96210603.
RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
cancer deletion region and demonstrate distinct expression
patterns.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR EMBL; U28369; G974284; -.
DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 749 AA; 83121 MW; 6138C5D5 CRC32;

Query Match 92.4%; Score 73; DB 4; Length 749;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 530 RDPYCAWD 537
QY 475 RDPYCGWD 482

RESULT 6
ID Q93018 PRELIMINARY; PRT; 750 AA.
AC Q93018;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
```


RP SEQUENCE FROM N.A.
RA DANTE M., WAMLEY P.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN (2)
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN (3)
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U73167; G2880035; -;
DR PFAM: PF00047; IG: 1;
SQ SEQUENCE 750 AA; 83034 MW; CBC87E34 CRC32;

Query Match 92.4%; Score 73; DB 4; Length 750;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 531 RDPYCAWD 538
|||||:
Qy 475 RDPYCGWD 482

RESULT 7
ID O42236 PRELIMINARY; PRT; 751 AA.
AC O42236;

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COLLAPIN 3.
OS GALLUS GALLUS (CHICKEN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA FEINER L., KOPPEL A.M., KOBAYASHI H., RAPER J.A.;
RL NEURON 19-0-0(1997).
DR EMBL: AF022946; G2522204; -;
SQ SEQUENCE 751 AA; 85432 MW; 0B9BA2F7 CRC32;

Query Match 92.4%; Score 73; DB 13; Length 751;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 528 RDPYCAWD 535
|||||:
Qy 475 RDPYCGWD 482

RESULT 8
ID Q99985 PRELIMINARY; PRT; 751 AA.
AC Q99985;

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA YAMADA T., ENDO R., GOTOH M., HIROHASHI S.;
RT "Identification of semaphorin E as a non-MDR drug resistance gene of human cancers."
RL PROC. NATL. ACAD. SCI. U.S.A. 94:14716-14718(1997).
DR EMBL: AB000220; D1033360; -;
SQ SEQUENCE 751 AA; 85207 MW; F89FB934 CRC32;

Query Match 92.4%; Score 73; DB 4; Length 751;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 528 RDPYCAWD 535
|||||:
Qy 475 RDPYCGWD 482

RESULT 9
ID Q62181 PRELIMINARY; PRT; 751 AA.
AC Q62181;

DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E PRECURSOR (SEM E).
GN SEMAE OR SEME.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCUTOGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI:
RX MEDLINE: 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).

CC -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC MAXIMUM EXPRESSION BETWEEN DAYS 10-12 WITH MODERATE LEVELS FROM
CC DAY 13 UNTIL BIRTH.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL: X85994; G854332; -;
DR MGD: MGI:107557; SEMAE.
DR PFAM: PF00047; IG: 1.
KW SIGNAL: IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KW DEVELOPMENTAL PROTEIN.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 751 SEMAPHORIN E.
FT DOMAIN 587 649 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 751 AA; 85259 MW; B28D6CFC CRC32;

Query Match 92.4%; Score 73; DB 11; Length 751;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 528 RDPYCAWD 535
|||||:
Qy 475 RDPYCGWD 482

RESULT 10
ID Q13372 PRELIMINARY; PRT; 753 AA.
AC Q13372;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN III FAMILY HOMOLOG.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN:
RX MEDLINE: 96230324.
RA XIANG R.H., HENSEL C.H., GARCIA D.K., CARLSON H.C., KOK K., DALY M.C.,
RA KREACHER K., DEN BERG A., VELDHOUIS P., BUYS C.H., NAYLOR S.L.;
RT "Isolation of the human semaphorin iii/F gene (SEMA3F) at chromosome
3p21, a region deleted in lung cancer.";
RL GENOMICS 32:39-48(1996).
DR EMBL: U38276; G1061351; -;
DR PFAM: PF00047; IG: 1.
SQ SEQUENCE 753 AA; 84941 MW; BECBEBB0 CRC32;

Query Match 92.4%; Score 73; DB 4; Length 753;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 530 RDPYCAWD 537
QY 475 RDPYCGWD 482
|||||:|

RESULT 11
ID O88633 PRELIMINARY; PRT; 754 AA.
AC O88633;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV ISOFORM A.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ECKHARDT F., MEYERHANS A.;
RT "Molecular cloning and expression pattern of a murine semaphorin homologous to H-sena IV.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF080091; G3377768; -.
SQ SEQUENCE 754 AA; 85133 MW; EC05BDF3 CRC32;

Query Match 92.4%; Score 73; DB 11; Length 754;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 531 RDPYCAWD 538
QY 475 RDPYCGWD 482
|||||:|

RESULT 12
ID Q90663 PRELIMINARY; PRT; 761 AA.
AC Q90663;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COLLAPSIIN-2.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 95329269.
RA LUO Y., SHEPHERD I., LI J., RENZI M.J., CHANG S., RAPER J.A.;
RT "A family of molecules related to collapsin in the embryonic chick nervous system.";
RL NEURON 14:1131-1140(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94006554.
RA LUO Y., RAIBLE D., RAPER J.A.;
RT "Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones.";
RL CELL 75:217-227(1993).
DR EMBL: U28240; G886809; -.
DR PFAM: PF00047; Ig; 1.
SQ SEQUENCE 761 AA; 87300 MW; F0533E6C CRC32;

Query Match 92.4%; Score 73; DB 13; Length 761;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 536 RDPYCAWD 543
QY 475 RDPYCGWD 482
|||||:|

Query Match 92.4%; Score 73; DB 4; Length 771;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 531 RDPYCAWD 538
QY 475 RDPYCGWD 482
|||||:|

RESULT 13
ID Q14563 PRELIMINARY; PRT; 771 AA.
AC Q14563;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN-III.
GN HSEMA-III.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 94094332.
RA KOLODKIN A.L., MATTHES D.J., GOODMAN C.S.;
RT "The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules.";
RL CELL 75:1389-1399(1993).
RN [2]
RP SEQUENCE OF 1-37 FROM N.A.
RA WOESSNER J., MINX P., HINDS K., STROMMATT C.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 1-37 FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: L26081; G436560; -.
DR EMBL: AC004451; G2979605; -.
DR PFAM: PF00047; Ig; 1.
SQ SEQUENCE 771 AA; 88889 MW; 9EB1A137 CRC32;

Query Match 92.4%; Score 73; DB 4; Length 771;
Best Local Similarity 87.5%; Pred. No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 531 RDPYCAWD 538
QY 475 RDPYCGWD 482
|||||:|

RESULT 14
ID Q62180 PRELIMINARY; PRT; 772 AA.
AC Q62180; Q08665; Q52215;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN D PRECURSOR (SEMAPHORIN III) (COLLAPSIIN-1) (SEM D) (SEMA III).
GN SEMAD OR SEMD OR SEM III.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA TANIGUCHI M.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 107-772 FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 95267432.
RA MESSERSMITH E.K., LEONARDO E.D., SHATZ C.J., TESSIER-LAVIGNE M., GOODMAN C.S., KOLODKIN A.L.;

RT "Semaphorin III can function as a selective chemorepellent to pattern
RT sensory projections in the spinal cord.";
RL NEURON 14:949-959(1995)
CC -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC SELECTIVELY REPELS AXONS WHICH NORMALLY TERMINATE DORSALLY.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC MAXIMUM EXPRESSION BETWEEN DAYS 10-12 WITH MODERATE LEVELS FROM
CC DAY 13 UNTIL BIRTH.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; D85028; D1020554; -
DR EMBL; X85993; G854330; -
DR EMBL; L40484; G703190; -
DR MGD; MGI:107558; SEMAD.
DR PFAM; PF00047; ig; 1.
KW SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KW DEVELOPMENTAL PROTEIN.
FT SIGNAL 1 20 POTENTIAL
FT CHAIN 21 772 SEMAPHORIN D.
FT DOMAIN 593 636 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 772 AA; 88710 MW; B929B537 CRC32;

Query Match 92.4%; Score 73; DB 11; Length 772;
Best Local Similarity 87.5%; Pred.No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 531 RDPYCAWD 538
QY 475 RDPYCGWD 482
|||||:||

RESULT 15
ID Q90607 PRELIMINARY; PRT: 772 AA.
AC Q90607;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COLLAPSID.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=BRAIN;
RX MEDLINE; 94006554.
RA LUO Y., RAIBLE D., RAPER J.A.;
RT "Collapsin: a protein in brain that induces the collapse and
RT paralysis of neuronal growth cones."
RL CELL 75:1217-227(1993).
DR EMBL; U02528; G410079; -
DR PFAM; PF00047; ig; 1.
SQ SEQUENCE 772 AA; 88867 MW; BC6CB9E8 CRC32;

Query Match 92.4%; Score 73; DB 13; Length 772;
Best Local Similarity 87.5%; Pred.No. 4.83e-04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 531 RDPYCAWD 538
QY 475 RDPYCGWD 482
|||||:||

Search completed: Thu Jul 8 18:57:55 1999
Job time : 14 secs.

THIS PAGE BLANK (USPTO)

WQERH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:58:15 1999; MasPar time 5.04 Seconds
63.609 Million cell updates/sec
Tabular output not generated.

Title: >US-09-041-236-2
Description: (475-482) from US09041236.pep (25 of 45)
Perfect Score: 79
Sequence: 1 RDPYCGWD 8

Scoring table: PAM 150
Gap 11
Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.687; Variance 34.260; scale 0.691

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	79	100.0	724	2	C49423 semaphorin II precurs	4.52e-05
2	73	92.4	666	2	I58169 semaphorin III - mous	8.77e-04
3	73	92.4	748	2	I48744 semaphorin A - mouse	8.77e-04
4	73	92.4	749	2	G01856 semaphorin V - human	8.77e-04
5	73	92.4	751	2	I48748 semaphorin E - mouse	8.77e-04
6	73	92.4	753	2	G02173 semaphorin III family	8.77e-04
7	73	92.4	771	2	D49423 semaphorin III precur	8.77e-04
8	73	92.4	772	2	I48747 semaphorin D - mouse	8.77e-04
9	73	92.4	772	2	A49069 collapsin - chicken	8.77e-04
10	71	89.9	1074	2	JC5928 semaphorin F precursor	2.31e-03
11	70	88.6	834	2	S68498 M-sema F protein prec	3.73e-03
12	67	84.8	782	2	I48746 semaphorin C - mouse	1.54e-02
13	65	82.3	656	2	B49423 semaphorin I - fruit	3.91e-02
14	65	82.3	730	2	JH0798 fasciclin IV precursor	3.91e-02
15	61	77.2	760	2	I48745 semaphorin B - mouse	2.41e-01
16	61	77.2	1894	2	JC4980 plexin 1 precursor -	2.41e-01
17	61	77.2	1905	2	I54553 Plexin - African claw	2.41e-01
18	56	70.9	502	2	I30010 NADH dehydrogenase (u	2.15e+00
19	56	70.9	590	2	S34960 NADH dehydrogenase (u	2.15e+00
20	56	70.9	590	1	QOUTC5 NADH dehydrogenase (u	2.15e+00
21	56	70.9	1110	2	I59370 guanylate cyclase (EC	2.15e+00
22	56	70.9	1146	3	T02766 WI-1 adhesin - Ajello	2.15e+00
23	56	70.9	1157	2	A55152 PAS1 protein - yeast	2.15e+00

24	56	70.9	1630	2	J01168 polyprotein - turnip	2.15e+00
25	56	70.9	3163	1	JQ1895 genome polyprotein -	2.15e+00
26	55	69.6	2255	1	JQ1532 genome polyprotein -	3.30e+00
27	54	68.4	286	2	JE0417 aminoglycoside-N-acet	5.02e+00
28	54	68.4	383	2	S75240 hypothetical protein	5.02e+00
29	54	68.4	417	2	S25627 glucose-1-phosphatase	5.02e+00
30	54	68.4	481	2	A58429 I-kappa-B-related pro	5.02e+00
31	53	67.1	225	1	W2BEF1 gene 58 protein - equ	7.60e+00
32	53	67.1	353	2	JAO123 replicase - potato vi	7.60e+00
33	53	67.1	465	2	JCS184 glutamate synthase (G	7.60e+00
34	53	67.1	472	2	G65112 glutamate synthase (N	7.60e+00
35	52	65.8	711	2	A49423 semaphorin I precursor	7.60e+00
36	52	65.8	105	2	S44125 Ig lambda chain v reg	1.15e+01
37	52	65.8	113	1	HRTHBD hemerythrin - sipunc	1.15e+01
38	52	65.8	180	2	S76762 hypothetical protein	1.15e+01
39	52	65.8	196	1	K1XFAS thymidine kinase (EC	1.15e+01
40	52	65.8	272	2	T01480 hypothetical protein	1.15e+01
41	52	65.8	1872	2	JC4976 plexin 3 precursor -	1.15e+01
42	52	65.8	1884	2	JC4975 plexin 2 precursor -	1.15e+01
43	51	64.6	1004	2	A39611 probable GTP-binding	1.72e+01
44	51	64.6	1707	2	S77908 hypothetical protein	1.72e+01
45	51	64.6	1737	2	T00209 MEGF8 protein - human	1.72e+01

ALIGNMENTS

RESULT 1
ENTRY #type complete
TITLE semaphorin II precursor - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 04-Sep-1998
ACCESSIONS C49423
REFERENCE A49423
#authors Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
#journal Cell (1993) 75:1389-1399
#title The Semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules.

#accession C49423
#status preliminary; nucleic acid sequence not shown; not compared with conceptual translation
#molecule_type mRNA
#residues_type 1-724 #label KOL
#cross-references GB:L26083

GENETICS
#gene sema II
#cross-references FlyBase:FBgn0011260
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 724 #molecular-weight 82971 #checksum 6082

Query Match 100.0%; Score 79; DB 2; Length 724;
Best Local Similarity 100.0%; Pred. No. 4.52e-05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 537 RDPYCGWD 544

Qy 475 RDPYCGWD 482

RESULT 2
ENTRY #type fragment
TITLE semaphorin III - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 04-Sep-1998

ACCESSIONS I58169
REFERENCE I58169
#authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.; Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#journal Neuron (1995) 14:949-959
#title Semaphorin III can function as a selective chemorepellent to pattern sensory projections in the spinal cord.
#cross-references MUID:95267432

```
#accession I58169      preliminary: translated from GB/EMBL/DBJ
##status
##molecule_type mRNA
##residues 1-666 ##label RES
##cross-references GB:L40484; NID:g703189; PID:g703190
GENETICS
  gene SemAIII
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 666 #checksum 9654
  Query Match 92.4%; Score 73; DB 2; Length 666;
  Best Local Similarity 87.5%; Pred. No. 8.77e-04;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 425 RDPYCAWD 432
Qy 475 RDPYCGWD 482
  |||||:|
  #type complete
  #formal_name Mus musculus #common_name house mouse
  #sequence_revision 02-Jul-1996 #text_change
  04-Sep-1998
ACCESSIONS I48744
REFERENCE I48744
  #authors Puschel, A.W.; Adams, R.H.; Betz, H.
  #journal Neuron (1995) 14:941-948
  #title Murine semaphorin D/collapsin is a member of a diverse gene
  #family and creates domains inhibitory for axonal extension.
  #cross-references MUID:95267431
  #accession I48748
  ##status preliminary: translated from GB/EMBL/DBJ
  ##molecule_type mRNA
  ##residues 1-751 ##label RES
  ##cross-references EMBL:X85994; NID:g854331; PID:g854332
GENETICS
  gene semE
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 751 #molecular-weight 85259 #checksum 8961
  Query Match 92.4%; Score 73; DB 2; Length 751;
  Best Local Similarity 87.5%; Pred. No. 8.77e-04;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 528 RDPYCAWD 535
Qy 475 RDPYCGWD 482
  |||||:|
  #type complete
  #formal_name Homo sapiens #common_name man
  #sequence_revision 06-Jun-1997 #text_change
  04-Sep-1998
ACCESSIONS G02173
REFERENCE G09275
  #authors Naylor, S
  #submission submitted to the EMBL Data Library, October 1995
  #accession G02173
  ##status preliminary: translated from GB/EMBL/DBJ
  ##molecule_type mRNA
  ##residues 1-753 ##label NAY
  ##cross-references EMBL:U38276; NID:g1061350; PID:g1061351
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 753 #molecular-weight 84941 #checksum 5681
  Query Match 92.4%; Score 73; DB 2; Length 753;
  Best Local Similarity 87.5%; Pred. No. 8.77e-04;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 530 RDPYCAWD 537
Qy 475 RDPYCGWD 482
  |||||:|
  #type complete
  #formal_name Homo sapiens #common_name man
  #sequence_revision 06-Jan-1995 #text_change
  04-Sep-1998
ACCESSIONS D49423
REFERENCE D49423
  #authors Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
  #type complete
  #formal_name Homo sapiens #common_name man
  #sequence_revision 06-Jan-1995 #text_change
  04-Sep-1998
ACCESSIONS D49423
REFERENCE D49423
  #authors
```

```
#accession I58169      preliminary: translated from GB/EMBL/DBJ
##status
##molecule_type mRNA
##residues 1-666 ##label RES
##cross-references GB:L40484; NID:g703189; PID:g703190
GENETICS
  gene SemAIII
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 666 #checksum 9654
  Query Match 92.4%; Score 73; DB 2; Length 666;
  Best Local Similarity 87.5%; Pred. No. 8.77e-04;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 425 RDPYCAWD 432
Qy 475 RDPYCGWD 482
  |||||:|
  #type complete
  #formal_name Mus musculus #common_name house mouse
  #sequence_revision 02-Jul-1996 #text_change
  04-Sep-1998
ACCESSIONS I48744
REFERENCE I48744
  #authors Puschel, A.W.; Adams, R.H.; Betz, H.
  #journal Neuron (1995) 14:941-948
  #title Murine semaphorin D/collapsin is a member of a diverse gene
  #family and creates domains inhibitory for axonal extension.
  #cross-references MUID:95267431
  #accession I48744
  ##status preliminary: translated from GB/EMBL/DBJ
  ##molecule_type mRNA
  ##residues 1-748 ##label RES
  ##cross-references EMBL:X85990; NID:g854323; PID:g854324
GENETICS
  gene semA
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 748 #molecular-weight 82894 #checksum 9017
  Query Match 92.4%; Score 73; DB 2; Length 748;
  Best Local Similarity 87.5%; Pred. No. 8.77e-04;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 529 RDPYCAWD 536
Qy 475 RDPYCGWD 482
  |||||:|
  #type complete
  #formal_name Homo sapiens #common_name man
  #sequence_revision 06-Jun-1997 #text_change
  04-Sep-1998
ACCESSIONS G01856
REFERENCE G08634
  #authors Sekido, Y.
  #submission submitted to the EMBL Data Library, June 1995
  #accession G01856
  ##status preliminary: translated from GB/EMBL/DBJ
  ##molecule_type mRNA
  ##residues 1-749 ##label SEK
  ##cross-references EMBL:U28369; NID:g974283; PID:g974284
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 749 #molecular-weight 83121 #checksum 2747
  Query Match 92.4%; Score 73; DB 2; Length 749;
  Best Local Similarity 87.5%; Pred. No. 8.77e-04;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
#journal      Cell (1993) 75:1389-1399
#title        The Semaphorin genes encode a family of transmembrane and
#accession    D49433
#status       preliminary; nucleic acid sequence not shown
##molecule_type mRNA
##residues    1-771 ##label KOL
##cross-references GB:L26081; NID:g799328; PID:g436560
GENETICS
#gene         GDB:SEMA1
##cross-references GDB:283448
CLASSIFICATION #superfamily semaphorin
SUMMARY       #length 771 #molecular-weight 88889 #checksum 6249
Query Match   92.4%; Score 73; DB 2; Length 771;
Best Local Similarity 87.5%; Pred. No. 8.77e-04;
Matches       7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 531 RDPYCGWD 538
QY 475 RDPYCGWD 482
|||||:|

RESULT      8
ENTRY       I48747 #type complete
TITLE       semaphorin D - mouse
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
ACCESSIONS  I48747
REFERENCE    I48744
#authors     Puschel, A.W.; Adams, R.H.; Betz, H.
#journal     Neuron (1993) 14:941-948
#title       Murine semaphorin D/collapsin is a member of a diverse gene
#comment     family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession   I48747
##status     preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues   1-772 ##label RES
##cross-references EMBL:X85993; NID:g854329; PID:g854330
GENETICS
#gene        semD
#superfamily semaphorin
CLASSIFICATION #length 772 #molecular-weight 88710 #checksum 1776
SUMMARY
Query Match   92.4%; Score 73; DB 2; Length 772;
Best Local Similarity 87.5%; Pred. No. 8.77e-04;
Matches       7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 531 RDPYCGWD 538
QY 475 RDPYCGWD 482
|||||:|

RESULT      9
ENTRY       A49069 #type complete
TITLE       collapsin - chicken
ORGANISM    #formal_name Gallus gallus #common_name chicken
DATE        07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
ACCESSIONS  A49069
REFERENCE    A49069
#authors     Luo, Y.; Raible, D.; Raper, J.A.
#journal     Cell (1993) 75:217-227
#title       Collapsin: a protein in brain that induces the collapse and
#comment     paralysis of neuronal growth cones.
#accession   A49069
##status     preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues   1-772 ##label LUO
##cross-references GB:U02528; NID:g410078; PID:g410079
CLASSIFICATION #superfamily semaphorin
```

```
SUMMARY       #length 772 #molecular-weight 88867 #checksum 9712
Query Match   92.4%; Score 73; DB 2; Length 772;
Best Local Similarity 87.5%; Pred. No. 8.77e-04;
Matches       7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 531 RDPYCGWD 538
QY 475 RDPYCGWD 482
|||||:|

RESULT      10
ENTRY       JC5928 #type complete
TITLE       semaphorin F precursor - human
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        10-Apr-1998 #sequence_revision 08-May-1998 #text_change
ACCESSIONS  JC5928
REFERENCE    JC5928
#authors     Simmons, A.D.; Puschel, A.W.; McPherson, J.D.; Overhauser,
#journal     J.; Lovett, M.
#title       Biochem. Biophys. Res. Commun. (1998) 242:685-691
#comment     Molecular cloning and mapping of human semaphorin F from the
#comment     Cri-du-chat candidate interval.
#accession   JC5928
##status     nucleic acid sequence not shown
##molecule_type mRNA
##residues   1-1074 ##label SIM
##cross-references GB:U52840; NID:g2772583; PID:g2772584
#comment     This protein disrupts normal brain development and leads to some of
#comment     the features of Cri-du-chat.
GENETICS
#gene        semaf
#superfamily thrombospondin type 1 repeat homology
CLASSIFICATION #domain signal sequence #status predicted #label SIG\
FEATURE       1-20 #domain semaphorin #status predicted #label SEM\
50-533 #domain thrombospondin type 1 repeat homology #label
840-896 THR3\
971-993 #domain transmembrane #status predicted #label TMM
SUMMARY       #length 1074 #molecular-weight 120570 #checksum 2910
Query Match   89.9%; Score 71; DB 2; Length 1074;
Best Local Similarity 87.5%; Pred. No. 2.31e-03;
Matches       7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 500 QDPYCGWD 507
QY 475 RDPYCGWD 482
|||||:|

RESULT      11
ENTRY       S66498 #type complete
TITLE       M-sema F protein precursor - mouse
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change
ACCESSIONS  S66498
REFERENCE    S66498
#authors     Inagaki, S.; Furuyama, T.; Iwahashi, Y.
#journal     FEBS Lett. (1995) 370:269-272
#title       Identification of a member of mouse semaphorin family.
#comment     #cross-references MUID:95385809
#accession   S66498
##status     preliminary
##molecule_type mRNA
##residues   1-834 ##label INA
##cross-references EMBL:S79463; NID:gl110598; PID:gl110599
FEATURE       1-21 #domain signal sequence #status predicted #label SIG\
22-834 #product M-sema F protein #status predicted #label MAT
SUMMARY       #length 834 #molecular-weight 92556 #checksum 7189
```

```
Query Match      88.6%; Score 70; DB 2; Length 834;
Best Local Similarity 75.0%; Pred. No. 3.73e-03;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 513 RDPYCAWN 520
QY 475 RDPYCGWD 482
|||||:|

RESULT 12
ENTRY      I48746      #type fragment
TITLE      semaphorin C - mouse (fragment)
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE       02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
28-Feb-1997
ACCESSIONS I48746
REFERENCE   I48744
#authors   Puschel, A.W.; Adams, R.H.; Betz, H.
#journal   Neuron (1995) 14:941-948
#title     Murine semaphorin D/collapsin is a member of a diverse gene
           family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48746      preliminary; translated from GB/EMBL/DBBJ
#molecule_type mRNA
#status     1-782 #label RES
#residues   1-782
#cross-references EMBL:X85992; NID:g8543327; PID:g8543328
GENETICS
#gene       semC
SUMMARY     #length 782 #checksum 1571

Query Match      84.8%; Score 67; DB 2; Length 782;
Best Local Similarity 85.7%; Pred. No. 1.54e-02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 485 RDPYCAW 491
QY 475 RDPYCGW 481
|||||:|

RESULT 13
ENTRY      B49423      #type fragment
TITLE      semaphorin I - fruit fly (Drosophila melanogaster) (fragment)
ORGANISM   #formal_name Drosophila melanogaster
DATE       06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
16-Feb-1997
ACCESSIONS B49423
REFERENCE   A49423
#authors   Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
#journal   Cell (1993) 75:1389-1399
#title     The semaphorin genes encode a family of transmembrane and
           secreted growth cone guidance molecules.
#accession B49423
#status     preliminary; nucleic acid sequence not shown; not
           compared with conceptual translation
#molecule_type mRNA
#residues   1-656 #label KOL
#cross-references GB:I26082
GENETICS
#gene       semaI
SUMMARY     #length 656 #checksum 3860

Query Match      82.3%; Score 65; DB 2; Length 656;
Best Local Similarity 75.0%; Pred. No. 3.91e-02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 414 ODPYCAWD 421
QY 475 RDPYCGWD 482
|||||:|

Query Match      82.3%; Score 65; DB 2; Length 730;
Best Local Similarity 75.0%; Pred. No. 3.91e-02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 508 QDPYCAWD 515
QY 475 RDPYCGWD 482
|||||:|

RESULT 14
ENTRY      JH0798      #type complete
TITLE      fasciclin IV precursor - American bird grasshopper
ORGANISM   #formal_name Schistocerca americana #common_name American
           bird grasshopper
DATE       30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
09-Sep-1997
ACCESSIONS JH0798
REFERENCE   JH0798
#authors   Kolodkin, A.; Matthes, D.J.; O'Connor, T.P.; Patel, N.H.;
           Admon, A.; Bentley, D.; Goodman, C.S.
#journal   Neuron (1992) 9:831-845
#title     Fasciclin IV: Sequence, expression, and function during
           growth cone guidance in the grasshopper embryo.
#accession JH0798
#molecule_type mRNA
#residues   1-730 #label KOL
#cross-references GB:L00709; NID:g160844; PID:g160845
#experimental_source embryo
COMMENT     This protein plays a role in growth cone guidance in the developing
           central nervous system.
KEYWORDS    glycoprotein; transmembrane protein
FEATURES
1-22        #domain signal sequence #status predicted #label SIG\
23-730      #product fasciclin IV #status predicted #label MAT\
23-627      #domain extracellular #status predicted #label EXT\
628-652     #domain transmembrane #status predicted #label TM\
653-730     #domain intracellular #status predicted #label INT\
44,71,163,267,360,
539         #binding_site carbohydrate (Asn) (covalent) #status
           predicted
SUMMARY     #length 730 #molecular-weight 81214 #checksum 5881

Query Match      82.3%; Score 65; DB 2; Length 730;
Best Local Similarity 75.0%; Pred. No. 3.91e-02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 508 QDPYCAWD 515
QY 475 RDPYCGWD 482
|||||:|

RESULT 15
ENTRY      I48745      #type complete
TITLE      semaphorin B - mouse
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE       02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
28-Feb-1997
ACCESSIONS I48745
REFERENCE   I48744
#authors   Puschel, A.W.; Adams, R.H.; Betz, H.
#journal   Neuron (1995) 14:941-948
#title     Murine semaphorin D/collapsin is a member of a diverse gene
           family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48745      preliminary; translated from GB/EMBL/DBBJ
#status     1-760 #label RES
#molecule_type mRNA
#residues   1-760
#cross-references EMBL:X85991; NID:g854325; PID:g854326
GENETICS
#gene       semaB
SUMMARY     #length 760 #molecular-weight 83458 #checksum 2188

Query Match      77.2%; Score 61; DB 2; Length 760;
Best Local Similarity 75.0%; Pred. No. 2.41e-01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 510 RDPHCAWD 517
QY 475 RDPYCGWD 482
|||||:|
```


Mon Jul 12 06:28:52 1999

US-09-041-236-2-25.rpr

Page 5

Search completed: Thu Jul 8 18:58:24 1999
Job time : 9 secs.

THIS PAGE BLANK (USPTO)

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Thu Jul 8 18:40:56 1999; MasPar time 4.96 Seconds
Tabular output not generated. 72.697 Million cell updates/sec

Title: >US-09-041-236-2
Description: (237-245) from US09041236.ppt (17 of 45)
Perfect Score: 53
Sequence: 1 DQGGESSLS 9

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 20.028; Variance 19.976; scale 1.003

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	53	100.0	653	2	T03102 semaphorin homolog A3	8.53e+03
2	48	90.6	966	1	PHPOAG starch phosphorylase	2.04e-01
3	41	77.4	93	2	S42743 hypothetical protein	1.25e+01
4	41	77.4	295	2	J01775 Sal19R protein - vacc	1.25e+01
5	41	77.4	403	2	E42521 A39R protein - vaccin	1.25e+01
6	41	77.4	441	2	S2921 hypothetical protein	1.25e+01
7	41	77.4	534	1	P1WL41 L1 protein - human pa	1.25e+01
8	41	77.4	796	2	S46693 hypothetical protein	1.25e+01
9	41	77.4	835	2	A49891 outer membrane usher	1.25e+01
10	41	77.4	2137	1	SJHUB spectrin beta chain -	1.25e+01
11	41	77.4	2411	2	A48299 tyrosine kinase subst	1.25e+01
12	40	75.5	186	2	PQ0200 glycinin G5 - soybea	2.17e+01
13	40	75.5	274	1	MNVNM nonstructural protein	2.17e+01
14	40	75.5	274	1	MNVNJ nonstructural protein	2.17e+01
15	40	75.5	390	2	D64903 hypothetical protein	2.17e+01
16	40	75.5	565	2	T01116 carrot B2 protein hom	2.17e+01
17	40	75.5	588	2	F70971 hypothetical glycine	2.17e+01
18	40	75.5	673	2	A45456 NADH dehydrogenase (u	2.17e+01
19	40	75.5	675	2	G71640 NADH dehydrogenase I	2.17e+01
20	40	75.5	675	2	S53832 NADH dehydrogenase (u	2.17e+01
21	40	75.5	691	2	S78135 NADH dehydrogenase (u	2.17e+01
22	40	75.5	727	2	S17854 NADH dehydrogenase (u	2.17e+01
23	40	75.5	727	2	A33552 NADH dehydrogenase (u	2.17e+01

24	40	75.5	738	2	S52737 NADH dehydrogenase (u	2.17e+01
25	40	75.5	744	2	S59926 NADH dehydrogenase (u	2.17e+01
26	40	75.5	744	2	S17664 NADH dehydrogenase (u	2.17e+01
27	40	75.5	880	2	T02245 hypothetical protein	2.17e+01
28	39	73.6	191	2	PQ0810 glycinin A3B4 (plasm	3.73e+01
29	39	73.6	236	2	PQ0806 glycinin A3B4 - soybe	3.73e+01
30	39	73.6	238	2	PQ0809 glycinin A3B4 (plasm	3.73e+01
31	39	73.6	243	2	PQ0807 glycinin A3B4 - soybe	3.73e+01
32	39	73.6	336	2	C16159 benzoate 1,2-dioxyge	3.73e+01
33	39	73.6	382	2	B65160 mannitol-1-phosphate	3.73e+01
34	39	73.6	434	1	S49457 pyrimidine-nucleoside	3.73e+01
35	39	73.6	441	2	S74626 erythroid ankyrin - S	3.73e+01
36	39	73.6	452	2	S58938 fructose-2,6-bisphosp	3.73e+01
37	39	73.6	516	1	FWSRG3 glycinin G5 precursor	3.73e+01
38	39	73.6	540	2	T03309 probable terminase AA	3.73e+01
39	39	73.6	677	2	E70722 hypothetical protein	3.73e+01
40	39	73.6	1377	2	C70148 DNA-directed RNA poly	3.73e+01
41	39	73.6	1410	2	A57013 early endosome antige	3.73e+01
42	38	71.7	47	2	S62326 cystatin C, parocid 9	6.33e+01
43	38	71.7	134	2	C57233 complexin II - rat	6.33e+01
44	38	71.7	849	2	S64732 scaffold attachment f	6.33e+01
45	38	71.7	1466	2	S65741 1-phosphatidylinosito	6.33e+01

ALIGNMENTS

RESULT 1
ENTRY T03102 #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS T03102
REFERENCE Z14840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-653 ##label ENS
##cross-references EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY #length 653 #molecular-weight 73645 #checksum 5501
Query Match 100.0%; Score 53; DB 2; Length 653;
Best Local Similarity 100.0%; Pred. No. 8.53e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 291 DQGGESSLS 299
Oy 237 DQGGESSLS 245
|||||||

RESULT 2
ENTRY PHPOAG #type complete
TITLE starch phosphorylase (EC 2.4.1.1) precursor - potato
ALTERNATE_NAMES alpha-glucan phosphorylase
ORGANISM #formal_name Solanum tuberosum #common_name potato
DATE 04-Dec-1986 #sequence_revision 30-Sep-1990 #text_change 05-Dec-1998
ACCESSIONS JU0130; A00574; PQ0139; S15531; S12033
REFERENCE A91915
#authors Nakano, K.; Mori, H.; Fukui, T.
#journal J. Biochem. (1989) 106:691-695
#title Molecular cloning of cDNA encoding potato amyloplast
#cross-references MIM:90110071
#accession JU0130
##molecule_type mRNA
##residues 1-966 ##label NA1
##cross-references GB:D00520; NID:g217998; PID:d1000861; PID:g217999
REFERENCE A92591

```
#authors      Nakano, K.; Fukui, T.
#journal      J. Biol. Chem. (1986) 261:8230-8236
#title        The complete amino acid sequence of potato alpha-glucan
               phosphorylase.
#cross-references MUID:86250715
#accession    A00574
               ##molecule_type protein
               ##residues      51-966 ##label NAK
REFERENCE     PQ0139
#authors      Brisson, N.; Giroux, H.; Zollinger, M.; Camirand, A.; Simard,
               C.
#journal      Plant Cell (1989) 1:559-566
#title        Maturation and subcellular compartmentation of potato starch
               phosphorylase.
#cross-references MUID:92404721
#accession    PQ0139
               ##molecule_type mRNA
               ##residues      1-130 ##label BRI
               ##experimental_source tuber, cv. Kennebec
REFERENCE     S15531
#authors      Brisson, N.
#submission   Submitted to the EMBL Data Library, April 1990
#accession    S15531
               ##molecule_type mRNA
               ##residues      1-158, 'D', 160-966 ##label BR2
               ##cross-references EMBL:X52385; NID:G21578; PID:G21579
REFERENCE     S12033
#authors      Camirand, A.; St-Pierre, B.; Marneau, C.; Brisson, N.
#journal      Mol. Gen. Genet. (1990) 224:33-39
#title        Occurrence of a copia-like transposable element in one of the
               introns of the potato starch phosphorylase gene.
#cross-references MUID:91117174
#accession    S12033
               ##molecule_type mRNA
               ##residues      416-595 ##label CAM
               ##cross-references EMBL:X52385
COMMENT       Phosphorylase, an important allosteric enzyme in carbohydrate
               metabolism, catalyzes the formation of glucose 1-phosphate from
               polyglucose. Enzymes from different sources differ in their
               regulatory mechanisms and in their natural substrates. However,
               all known phosphorylases share catalytic and structural
               properties; the binding site for pyridoxal 5'-phosphate is highly
               conserved.
CLASSIFICATION #superfamily phosphorylase
KEYWORDS       allosteric regulation; carbohydrate metabolism;
               glycosyltransferase; hexosyltransferase; phosphoprotein;
               pyridoxal phosphate
FEATURE        1-50
               #domain transit peptide (amyloplast) #status predicted
               #label TNP\
51-966         #product phosphorylase #status experimental #label MAT\
812            #binding_site pyridoxal phosphate (Lys) (covalent)
               #status predicted
SUMMARY        #length 966 #molecular-weight 109505 #checksum 4821
               Query Match      90.6%; Score 48; DB 1; Length 966;
               Best Local Similarity 77.8%; Pred. No. 2.04e-01;
               Matches      7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db      62 EQGESDLS 70
        :|||||
Qy      237 DOGESLS 245

RESULT      3
ENTRY       S42743      #type fragment
TITLE       hypothetical protein - Erwinia chrysanthemi (fragment)
ORGANISM    07-Oct-1994 #formal_name Erwinia chrysanthemi
DATE        09-Sep-1997 #sequence_revision 26-May-1995 #text_change
ACCESSIONS  S42743
REFERENCE    S34943
#authors     Moulard, M.; Condemine, G.; Robert-Baudouy, J.
```

```
#journal      Mol. Microbiol. (1993) 8:685-695
#title        Characterization of the nucM gene coding for a nuclease of
               the phytopathogenic bacterium Erwinia chrysanthemi.
#accession    S42743
               ##status      preliminary; translation not shown
               ##molecule_type DNA
               ##residues      1-93 ##label MOU
               ##cross-references EMBL:X73255; NID:G403001; PID:G403003
SUMMARY      #length 93 #checksum 5663
               Query Match      77.4%; Score 41; DB 2; Length 93;
               Best Local Similarity 75.0%; Pred. No. 1.25e-01;
               Matches      6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db      15 QGGTDL 22
        :|||||
Qy      238 QGGESLS 245

RESULT      4
ENTRY       JQ1775      #type complete
TITLE       Sall9R protein - vaccinia virus (strain WR)
ORGANISM    30-Sep-1993 #formal_name vaccinia virus
DATE        09-Sep-1997 #sequence_revision 30-Sep-1993 #text_change
ACCESSIONS  JQ1775
REFERENCE    JQ1767
#authors     Smith, G.L.; Chan, Y.S.; Howard, S.T.
#journal      J. Gen. Virol. (1991) 72:1349-1376
#title        Nucleotide sequence of 42kbp of vaccinia virus strain WR from
               near the right inverted terminal repeat.
#accession    JQ1775
               ##molecule_type DNA
               ##residues      1-295 ##label SMI
               ##cross-references DDBJ:D11079; NID:G22717; PID:G1002287;
               ##checksum 7539
SUMMARY      #length 295 #molecular-weight 33699 #checksum 7539
               Query Match      77.4%; Score 41; DB 2; Length 295;
               Best Local Similarity 77.8%; Pred. No. 1.25e-01;
               Matches      7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db      236 DEGGPSSL 244
        :|||||
Qy      237 DOGESLS 245

RESULT      5
ENTRY       E42521      #type complete
TITLE       A39R protein - vaccinia virus (strain Copenhagen)
ORGANISM    host Homo sapiens (man)
DATE        09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change
ACCESSIONS  E42521
REFERENCE    A31172
#authors     Johnson, G.P.
#submission   Submitted to GenBank, June 1990
#accession    E42521
               ##status      preliminary
               ##molecule_type DNA
               ##residues      1-403 ##label JOH
               ##checksum 8167
SUMMARY      #length 403 #molecular-weight 45741 #checksum 8167
               Query Match      77.4%; Score 41; DB 2; Length 403;
               Best Local Similarity 77.8%; Pred. No. 1.25e-01;
               Matches      7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db      198 DEGGPSSL 206
        :|||||
Qy      237 DOGESLS 245

RESULT      6
```

```

S29921      #type complete
TITLE       hypothetical protein 15 - vaccinia virus
ORGANISM    #formal_name vaccinia virus
DATE        20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
           09-Sep-1997

ACCESSIONS  S29921
REFERENCE    S29907
#authors    Amegadzie, B.Y.
#submission submitted to the EMBL Data Library, January 1991
#accession  S29921
#status     preliminary
#molecule_type DNA
#residues   1-441 #label AME
#cross-references EMBL:X57318; NID:g62239; PID:g62254
SUMMARY     #length 441 #molecular-weight 50185 #checksum 6034

Query Match      77.4%; Score 41; DB 2; Length 441;
Best Local Similarity 77.8%; Pred. NO. 1.25e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 236 DEGGPSSL 244
      ||| ||||
Qy 237 DQGESL 245

RESULT      7
ENTRY       PIWL41      #type complete
TITLE       L1 protein - human papillomavirus type 41
ORGANISM    #formal_name human papillomavirus type 41
#note       host Homo sapiens (man)
DATE        30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
           27-Jan-1995

ACCESSIONS  H43550
REFERENCE    A43550
#authors    Hirt, L.; Hirsch-Behnman, A.; De Villiers, E.M.
#journal    Virus Res. (1990) 18:179-190
#title      Nucleotide sequence of human papillomavirus (HPV) type 41: an unusual HPV type without a typical E2 binding site consensus sequence.

#accession  H43550
#status     translation not shown
#molecule_type DNA
#residues   1-534 #label HIR
#cross-references EMBL:X56147
CLASSIFICATION #superfamily papillomavirus L1 protein
KEYWORDS      late protein
SUMMARY       #length 534 #molecular-weight 60949 #checksum 9283

Query Match      77.4%; Score 41; DB 1; Length 534;
Best Local Similarity 66.7%; Pred. NO. 1.25e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 163 DQGSDDL 171
      |||:| |
Qy 237 DQGESL 245

RESULT      8
ENTRY       S45693      #type complete
TITLE       hypothetical protein YHR204w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES #hypothetical protein H9998.22
ORGANISM        #formal_name Saccharomyces cerevisiae
DATE            28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change
           12-Dec-1997

ACCESSIONS  S45693
REFERENCE    S45671
#authors    Macri, C.
#submission submitted to the EMBL Data Library, February 1994
#description The sequence of S. cerevisiae cosmid 9998.
#accession  S45693
#molecule_type DNA
#residues   1-796 #label MAC

```

#journal J. Biol. Chem. (1991) 266:15154-15159
 #title A splice site mutation of the beta-spectrin gene causing exon skipping in hereditary elliptocytosis associated with a truncated beta-spectrin chain.

#cross-references MUID:91332035

#accession A93346

#molecule_type DNA

#residues 2002-2137 ##label GAL

#cross-references GB:J05500

REFERENCE JT0554

#authors Yoon, S.H.; Kentros, C.G.; Prchal, J.T.

#journal Gene (1990) 91:297-302

#title Identification of an unusual deletion within homologous repeats of human reticulocyte beta-spectrin and probable peptide polymorphism.

#cross-references MUID:91007291

#accession JT0554

#molecule_type mRNA

#residues 928-957, 'NY', 960-982, 'L', 984-1030, 'D', 1032-1755 ##label

YOO

#cross-references GB:M57948

REFERENCE A90733

#authors Winkermann, J.C.; Leto, T.L.; Watkins, P.C.; Eddy, R.; Shows, T.B.; Linnenbach, A.J.; Sahr, K.E.; Kathuria, N.; Marchesi, V.T.; Forget, B.G.

#journal Blood (1988) 72:328-334

#title Molecular cloning of the cDNA for human erythrocyte beta-spectrin.

#cross-references MUID:88269838

#accession A28777

#molecule_type mRNA

#residues 1334-1373, 'R', 1375-1432; 1909-2137 ##label WI2

#cross-references GB:J05500

#note authors Yoon et al. show His to be the predominant residue at position 1374 and Arg to be a minor allelic form or in error

REFERENCE A39885

#authors Prchal, J.T.; Morley, B.J.; Yoon, S.H.; Coetzer, T.L.; Palek, J.; Conboy, J.G.; Kan, Y.W.

#journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:7468-7472

#title Isolation and characterization of cDNA clones for human erythrocyte beta-spectrin.

#cross-references MUID:88041127

#accession A39885

#molecule_type mRNA

#residues 1209-1482 ##label PRC

#cross-references GB:M8054; MID:g338333; PID:g338334

#experimental_source reticulocyte

REFERENCE A23659

#authors Winkelmann, J.C.; Costa, F.F.; Linzie, B.L.; Forget, B.G.

#journal J. Biol. Chem. (1990) 265:20449-20454

#title Beta spectrin in human skeletal muscle. Tissue-specific differential processing of 3' beta spectrin pre-mRNA generates a beta spectrin isoform with a unique carboxyl terminus.

#cross-references MUID:91056094

#accession B23659

#molecule_type mRNA

#residues 2105-2137 ##label WI3

#cross-references GB:M37885

REFERENCE A42872

#authors Speicher, D.W.; Weglarz, L.; DeSilva, T.M.

#journal J. Biol. Chem. (1992) 267:14775-14782

#title Properties of human red cell spectrin heterodimer (side-to-side) assembly and identification of an essential nucleation site.

#cross-references MUID:92340516

#accession B42872

#molecule_type protein

#residues 47-56; 293-302; 1837-1846 ##label SP2

REFERENCE A93341

#authors Speicher, D.W.; Marchesi, V.T.

#journal Nature (1984) 311:177-180

#title Erythrocyte spectrin is comprised of many homologous triple helical segments.

#cross-references MUID:84295638

#accession B27016

#molecule_type protein

#residues 292-324, 'X', 326-329, 'Y', 331-332; 434-532; 718-734, 'V', 736-773, 'X', 775-777; 1036-1072, 'X', 1074-1077; 1386-1439, 'X', 1441-1443, 'X', 1445-1449; 1479-1509, 'X', 1511-1513; 1936-1988, 'XX', 1991-1992, 'X', 1994-1997 ##label SPE

#note the purified protein had a blocked amino end

COMMENT Spectrin is a major structural component of the erythrocyte membrane cytoskeleton and associates with ankyrin, actin, and band 4.1.

GENETICS

#gene GDB:SPTB

##cross-references GDB:119602; OMIM:182870

#map_position 14q23-14q23

CLASSIFICATION #superfamily spectrin beta chain; alpha-actinin actin-binding domain homology; spectrin/dystrophin repeat homology actin binding; cytoskeleton; duplication; erythrocyte; heterodimer; membrane protein

KEYWORDS

FEATURE

53-271

#domain alpha-actinin actin-binding domain homology #label ACT\

301-412 #domain spectrin/dystrophin repeat homology #label SN1\

421-526 #domain spectrin/dystrophin repeat homology #label SN2\

527-635 #domain spectrin/dystrophin repeat homology #label SN3\

636-741 #domain spectrin/dystrophin repeat homology #label SN4\

742-846 #domain spectrin/dystrophin repeat homology #label SN5\

847-952 #domain spectrin/dystrophin repeat homology #label SN6\

953-1059 #domain spectrin/dystrophin repeat homology #label SN7\

1060-1166 #domain spectrin/dystrophin repeat homology #label SN8\

1167-1272 #domain spectrin/dystrophin repeat homology #status atypical #label SN9\

1273-1377 #domain spectrin/dystrophin repeat homology #label S10\

1378-1476 #domain spectrin/dystrophin repeat homology #label S11\

1477-1582 #domain spectrin/dystrophin repeat homology #label S12\

1583-1688 #domain spectrin/dystrophin repeat homology #label S13\

1689-1795 #domain spectrin/dystrophin repeat homology #label S14\

1796-1901 #domain spectrin/dystrophin repeat homology #label S15\

1902-2007 #domain spectrin/dystrophin repeat homology #label S16\

2008-2118 #domain spectrin/dystrophin repeat homology #label S17\

SUMMARY #length 2137 #molecular-weight 246320 #checksum 7336

Query Match 77.4%; Score 41; DB 1; Length 2137;

Best Local Similarity 44.4%; Pred. No. 1.25e+01;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 1447 EGGDADLS 1455

: :: :: ::

QY 237 DQGGSSLS 245

RESULT 11

ENTRY

TITLE

A46299 #type complete

tyrosine kinase substrate disabled - fruit fly (Drosophila

melanogaster)

#formal_name Drosophila melanogaster

22-Sep-1993 #sequence_revision 25-Apr-1997 #text_change

01-Aug-1997

ACCESSIONS A46299

REFERENCE A46299

#authors Gertler, F.B.; Hill, K.K.; Clark, M.J.; Hoffmann, F.M.

#journal Genes Dev. (1993) 7:441-453

#title Dosage-sensitive modifiers of Drosophila abl tyrosine kinase

function: prospero, a regulator of axonal outgrowth, and

disabled, a novel tyrosine kinase substrate.

#cross-references MUID:93194063

#accession A46299

##status preliminary; not compared with conceptual translation

##molecule_type nucleic acid

##residues 1-2411 ##label GER

##cross-references GB:L08845

```
#note      sequence extracted from NCBI backbone (NCBIP:127163)
KEYWORDS   alternative splicing; phosphoprotein
SUMMARY    #length 2411 #molecular-weight 264046 #checksum 83

Query Match      77.4%; Score 41; DB 2; Length 2411;
Best Local Similarity 87.5%; Pred. No. 1.25e+01;
Matches          7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1306 DAGGESSL 1313
QY 237 DOGGESSL 244

RESULT 12
ENTRY   PQ0200      #type fragment
TITLE   glycinin G5 - soybean (fragment)
ORGANISM #formal_name Glycine max #common_name soybean
DATE    31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
        23-Mar-1995
ACCESSIONS PQ0200
REFERENCE   PQ0199
#authors   Scallion, B.; Thanh, V.H.; Floener, L.A.; Nielsen, N.C.
#journal   Theor. Appl. Genet. (1991) 70:510-519
#title     Identification and characterization of DNA clones encoding
          group-II glycinin subunits.
#accession PQ0200
#molecule_type DNA
##residues 1-186 #label SCA
##experimental_source embryo, strain CX635-1-1-1
##note     the authors translated the codon NAA for residue 93 as
          Gln and AGA for residue 95 as Lys
CLASSIFICATION #superfamily glycinin
KEYWORDS      seed; storage protein
FEATURE       16-186
SUMMARY      #product glycinin B4 chain (fragment) #status predicted
          #label MAT
          #length 186 #checksum 2491

Query Match      75.5%; Score 40; DB 2; Length 186;
Best Local Similarity 62.5%; Pred. No. 2.17e+01;
Matches          5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 127 EOGGERGL 134
QY 237 DOGGESSL 244

RESULT 13
ENTRY   MNVNM      #type complete
TITLE   nonstructural protein - vesicular stomatitis New Jersey virus
          (strain Missouri)
ORGANISM #formal_name vesicular stomatitis New Jersey virus
DATE    31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
        14-Nov-1997
ACCESSIONS A29143
REFERENCE   A29143
#authors   Rae, B.P.; Elliott, R.M.
#journal   J. Gen. Virol. (1986) 67:1351-1360
#title     Conservation of potential phosphorylation sites in the NS
          proteins of the New Jersey and Indiana serotypes of
          vesicular stomatitis virus.
#cross-references MUID:86253157
#accession A29143
#molecule_type mRNA
##residues 1-274 #label RAE
##cross-references EMBL:X04063; NID:961848; PID:961849
GENETICS
#gene      NS
CLASSIFICATION #superfamily vesicular stomatitis virus nonstructural protein
KEYWORDS      nonstructural protein
SUMMARY      #length 274 #molecular-weight 31300 #checksum 6939

Query Match      75.5%; Score 40; DB 1; Length 274;

#note      sequence extracted from NCBI backbone (NCBIP:127163)
KEYWORDS   alternative splicing; phosphoprotein
SUMMARY    #length 2411 #molecular-weight 264046 #checksum 83

Query Match      77.4%; Score 41; DB 2; Length 2411;
Best Local Similarity 87.5%; Pred. No. 1.25e+01;
Matches          7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1306 DAGGESSL 1313
QY 237 DOGGESSL 244

RESULT 12
ENTRY   PQ0200      #type fragment
TITLE   glycinin G5 - soybean (fragment)
ORGANISM #formal_name Glycine max #common_name soybean
DATE    31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
        23-Mar-1995
ACCESSIONS PQ0200
REFERENCE   PQ0199
#authors   Scallion, B.; Thanh, V.H.; Floener, L.A.; Nielsen, N.C.
#journal   Theor. Appl. Genet. (1991) 70:510-519
#title     Identification and characterization of DNA clones encoding
          group-II glycinin subunits.
#accession PQ0200
#molecule_type DNA
##residues 1-186 #label SCA
##experimental_source embryo, strain CX635-1-1-1
##note     the authors translated the codon NAA for residue 93 as
          Gln and AGA for residue 95 as Lys
CLASSIFICATION #superfamily glycinin
KEYWORDS      seed; storage protein
FEATURE       16-186
SUMMARY      #product glycinin B4 chain (fragment) #status predicted
          #label MAT
          #length 186 #checksum 2491

Query Match      75.5%; Score 40; DB 2; Length 186;
Best Local Similarity 62.5%; Pred. No. 2.17e+01;
Matches          5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 127 EOGGERGL 134
QY 237 DOGGESSL 244

RESULT 13
ENTRY   MNVNM      #type complete
TITLE   nonstructural protein - vesicular stomatitis New Jersey virus
          (strain Missouri)
ORGANISM #formal_name vesicular stomatitis New Jersey virus
DATE    31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
        14-Nov-1997
ACCESSIONS A29143
REFERENCE   A29143
#authors   Rae, B.P.; Elliott, R.M.
#journal   J. Gen. Virol. (1986) 67:1351-1360
#title     Conservation of potential phosphorylation sites in the NS
          proteins of the New Jersey and Indiana serotypes of
          vesicular stomatitis virus.
#cross-references MUID:86253157
#accession A29143
#molecule_type mRNA
##residues 1-274 #label RAE
##cross-references EMBL:X04063; NID:961848; PID:961849
GENETICS
#gene      NS
CLASSIFICATION #superfamily vesicular stomatitis virus nonstructural protein
KEYWORDS      nonstructural protein
SUMMARY      #length 274 #molecular-weight 31300 #checksum 6939

Query Match      75.5%; Score 40; DB 1; Length 274;

Best Local Similarity 66.7%; Pred. No. 2.17e+01;
Matches          6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 110 ENGGEKSL 118
QY 237 DOGGESSL 245

RESULT 14
ENTRY   MNVNVJ      #type complete
TITLE   nonstructural protein - vesicular stomatitis New Jersey virus
ORGANISM #formal_name vesicular stomatitis New Jersey virus
DATE    30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change
        14-Nov-1997
ACCESSIONS A04115
REFERENCE   A04115
#authors   Gill, D.S.; Banerjee, A.K.
#journal   J. Virol. (1985) 55:60-66
#title     Vesicular stomatitis virus NS proteins: structural similarity
          without extensive sequence homology.
#cross-references MUID:85237710
#accession A04115
#molecule_type mRNA
##residues 1-274 #label GIL
##cross-references GB:K03387; NID:9336047; PID:9336048
GENETICS
#gene      NS; M1
CLASSIFICATION #superfamily vesicular stomatitis virus nonstructural protein
KEYWORDS      nonstructural protein
SUMMARY      #length 274 #molecular-weight 31406 #checksum 8492

Query Match      75.5%; Score 40; DB 1; Length 274;
Best Local Similarity 66.7%; Pred. No. 2.17e+01;
Matches          6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 110 ENGGEKSL 118
QY 237 DOGGESSL 245

RESULT 15
ENTRY   D64903      #type complete
TITLE   hypothetical protein b1497 - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE    12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
        05-Dec-1998
ACCESSIONS D64903
REFERENCE   A64720
#authors   Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
          Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
          Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
          Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
          Y.
#journal   Science (1997) 277:1453-1462
#title     The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession D64903
##status   nucleic acid sequence not shown; translation not shown
#molecule_type DNA
##residues 1-390 #label BLAT
##cross-references GB:AE000247; GB:U000096; NID:91787773; PID:91787774;
          UMGF:b1497
#experimental_source strain K-12, substrain MG1655
GENETICS
#start_codon TTG
CLASSIFICATION #superfamily probable arylsulfatase regulatory protein atsb
SUMMARY      #length 390 #molecular-weight 45049 #checksum 3841

Query Match      75.5%; Score 40; DB 2; Length 390;
Best Local Similarity 62.5%; Pred. No. 2.17e+01;
Matches          5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 64 QGGEPTLA 71
```

Oy 238 OGGSSLS 245
||||::|:

Search completed: Thu Jul 8 18:41:05 1999
Job time : 9 secs.

Mouse Akt-2 cognat
IDUA - exons III t

ALIGNMENTS

Accession	Length	Accession	Length	Accession	Length
44	37.8	2.0	1599	1	T71252
45	37.6	2.0	4475	1	Q45995

RESULT 1

Location/Qualifiers

Database : N_Geneseq_34:*

```
FT mat_peptide /tag= b
FT 557.3016 /tag= c
FT
PN WO9717368-A1.
```

09-NOV-1995: US-556422.
 (DAND) DANA FARBER CANCER INST.
 BOUSLOGIOTIS V, FREEMAN GJ, HALL KT, NADLER LM, SCHULTZE JL;
 WPI: 97-280982/25.
 P-PSDB: W17658.
 Nucleic acid molecule encoding CD100 antigen - which stimulates
 leukocyte response, e.g. B cell aggregation, differentiation,
 survival and T cell proliferation
 Example 8; Page 81-83: 135pp: English.
 A cDNA molecule (760666) codes for mouse CD100 antigen (W17658) a
 novel leukocyte semaphorin-like protein that stimulates a leukocyte
 response, including B cell aggregation, B cell differentiation, B
 cell survival and/or T cell proliferation. It was isolated from a
 cDNA library prep. from murine T cells activated with anti-CD3
 and anti-CD28 antibodies using a probe corresponding to the
 semaphorin and Ig domains of the human CD100 sequence (see also
 760666). CD100 nucleic acids, antisense sequences, recombinant
 vectors, host cells expressing CD100, and CD100 polypeptides can
 be utilised in diagnostic and therapeutic methods involving
 modulation of B and T cell responses, neuron axonal growth and
 immune cell-nerve cell interaction.
 Sequence 4391 BP: 1009 A: 1180 C: 1105 G: 1097 T:

Query Match 3.0%; Score 57.8; DB 1; Length 4391;
Best Local Similarity 55.1%; Pred. NO. 0.00014;
Matches 135; Conservative 0; Mismatches 107; Indels 3; Gaps

[illegible]

```

RESULT 2
V31121
ID V31121 standard; cDNA; 2769 BP.
DT 02-SEP-1998 (first entry)
DE Human semaphorin encoding cDNA.
KW Human; semaphorin; diagnosis; nervous disease; immune disease;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 61..2646
/*tag= a
/product= "semaphorin"
J10155490-A.
PN PD 16-JUN-1998.
PF 27-NOV-1996; 332900.
PR 27-NOV-1996; JP-332900.
PA (SUMU ) SUMITOMO SEIYAKU KK.
DR WPI; 98-391044/34.
P-PSDB; W58540.
PT New human semaforin gene - useful in the diagnosis of nervous system
and immune disorders
PS Claim 2; Page 12-13; 15pp; Japanese.
CC The present sequence encodes human semaphorin (translated from the
Japanese specification as semaforin). Semaphorin has nerve extension
inhibiting activity. The semaphorin gene is useful for the diagnosis,
treatment and researches on nervous diseases and immune diseases.
SQ Sequence 2769 BP; 655 A; 780 C; 738 G; 596 T;

Query Match 3.0%; Score 57.8; DB 1; Length 2769;
Best Local Similarity 55.1%; Pred. No. 0.00012;
Matches 135; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

QY 605 AGGCTTACGATGACAGATCTACTACTCTTCCGAGAGAGACAATCCTGACAGAATCTCG 664
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 734 AGGGTGAGATGACAGAGTCTACTCTCTTTTACGGAGGTATCCGTGGAGTACGAATTCG 793

QY 665 AGGCTCTCTCAATGTGCCGTGTGCCAGTCTGTGCAGGGGACACAGGTGGGAAA 724
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 794 TCTCAAGTTGATGATCCGCGGAGTGGCCAGGTGTGCAAGGCGACGGCGCCCTGC 853

QY 725 GTTCACTGTGCTCCTCCAGTGGAAACACTTTTCTGAAAGCCATGCTGTGATGCACTGATG 784
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 854 GGACITTCGAAA---AAGTGGACCTCTTCTTAAAGGCCAGGCTGATCTGCTCCAGC 910

QY 785 CTGCCACCAACAAGAACTTCAACAGGCTGCAAGAGCTTCTCTGCTCCTGACCCCGCG 844
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 911 CAGACAGTGGCGCTGCTTCAACATCTTCAAGATGTGTTGTGCTGAGGCGCCCGGCC 970

QY 845 GCCAG 849
DB 971 TCAAG 975

RESULT 3
T60665
ID T60665 standard; cDNA; 4157 BP.
AC T60665;
DT 24-JUL-1997 (first entry)
DE Human CD100 antigen cDNA.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 88..2676
/*tag= a
signal_peptide 88..210
/*tag= b
FT mat_peptide 211..2673
/*tag= c

```

```

PN WO9717368-A1.
PD 15-MAY-1997.
PF 12-NOV-1996; UI8645.
PR 09-NOV-1995; US-556422.
PA (DAND ) DANA FARBER CANCER INST.
PI Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI; 97-280982/25.
P-PSDB; W17657.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
leukocyte response, e.g. B cell aggregation, differentiation,
survival and T cell proliferation
PS Claim 4; Page 65-69; 135pp; English.
CC A cDNA molecule (T60665) codes for human CD100 antigen (W17657) a
novel leukocyte semaphorin-like protein that stimulates a leukocyte
response, including B cell aggregation, B cell differentiation, B
cell survival and/or T cell proliferation. It was isolated by COS
cell expression cloning. Poly-A+ RNA from PHA-stimulated human T
cells was introduced into COS cells, and CD100-expressing cells
were selected by immunoselection and panning with CD100 monoclonal
antibody BD16. CD100 nucleic acids, antisense sequences,
recombinant vectors, host cells expressing CD100, and CD100
polypeptides can be utilised in diagnostic and therapeutic methods
involving modulation of B and T cell responses, neuron axonal
growth and immune cell-nerve cell interaction.
SQ Sequence 4157 BP; 960 A; 1094 C; 1058 G; 1045 T;

Query Match 2.8%; Score 53.8; DB 1; Length 4157;
Best Local Similarity 54.9%; Pred. No. 0.0011;
Matches 128; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

QY 611 ACGATGACAAGATCTACTACTTCTTCGAGAGGACAATCTCTGACAAGAATCTTGAGGCTC 670
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 767 AGGATGACAGGGTCTACTTCTTTCACGGAGGTGTCTGTGGAGTATGAGTTTGTGTCA 826

QY 671 CTCATATGTGTCCCGTGTGCCAGTTGTGCAGGGGGACAGGTGGGAAAATTTCAC 730
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 827 GGGTGTGATGATCCCGATAGCAAGAGTGTGCAAGGGGACAGGGCGGCTGAGGACCT 886

QY 731 TGTCACTCTCCAAGTGGAACTTTTCTGAAAGCCATGCTGGTATGAGTATGATGCTGCTCA 790
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 887 TGCAG---AGAAATGAGACCTCTTCTGAAAGCCCACTCATCTGCTCCGGCCAGACA 943

QY 791 CCAACAAGAAGTCTCAACAGGCTGCAAGACGTTCTTCTGCTCCCTGACCCCGCAGC 843
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 944 CGCGCTTGGTCTCAATGTGCTGGGATGCTCTGCTGCTCAGGTCCCGCGGC 996

RESULT 4
T72108
ID T72108 standard; cDNA to mRNA; 3524 BP.
AC T72108;
DT 13-FEB-1998 (first entry)
DE Human semaphorin 2 gene.
KW Semaphorin 2; central nerve extension; rat; human; inhibitor;
OS Homo sapiens.
FH Key Location/Qualifiers
FT 5'UTR 1..38
/*tag= a
FT CDS 39..2705
/*tag= b
FT 3'UTR 2706..3524
/*tag= c
WO9720928-A1.
PN PD 12-JUN-1997.
PF 02-DEC-1996; J03517.
PR 31-OCT-1996; JP-307205.
PR 06-DEC-1995; JP-345187.
PA (SUMU ) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI; 97-319775/29.
P-PSDB; W19857.

```

PT Semaphorin 2 and the gene encoding it - also inhibitors of its
PT action which can be used as promoters of central nerve regeneration
PS Claim 1; Page 72-74; 106pp; Japanese.
CC This sequence represents the human semaphorin 2 gene. The protein encoded
CC by this sequence, and the rat semaphorin 2 protein (see W19856) can be
CC used for screening possible candidates for activity as semaphorin 2
CC inhibitors. Proteins containing all or part of the semaphorin 2 sequence
CC are useful as central nerve extension inhibitors. Semaphorin 2 inhibitors
CC identified by screening can be used as promoters of central nerve
CC regeneration.
SQ Sequence 3524 BP; 516 A; 1212 C; 1204 G; 592 T;

Query Match 2.8%; Score 52.6; DB 1; Length 3524;
Best Local Similarity 44.1%; Pred. No. 0.002;
Matches 266; Conservative 0; Mismatches 334; Indels 3; Gaps 1;
Qy 321 GCTGCTGGCTGTGGCACCACCGCCGCGCCAGCTGCTGGAACCTGGTGAATGGCAC 380
Db 458 GCTCTTTGTGGGTTCACAGCCCTTCAACCCGGTGTGGCCCACTACAGCATAGACAC 517
Qy 381 TGTGTGGCACTTGGCGAGATGAGAGGTACGCCCTTTCAGCCGGGACGAGAACTCCCT 440
Db 518 COTGAGCCCGCTGGGAGACACATCAGCGGTATGGCCGCTGCCGCTACGACCCCAAGCA 577
Qy 441 GCTTCTGTTGAAGGGAGCGAGGTGTATTCACCATCCGGAGCAGGATACAAATGGGNA 500
Db 578 GCCCAATGTTGCCCTTCTCTGCGGGATGCTCTACAGCTACTGTACCGACTTCT 637
Qy 501 GATCCCTCGGTTCCGCGCATCCGGGGGAGAGTGAGTGATACACGATGATAGTGCAT 560
Db 638 AGCCATTGATGCTGTCATCTACCGCAGCTCGGGACAGGCCACCTCGCGCACCGTGAA 597
Qy 561 GCAGAACCCACAGTTTCATAAGCCACCACCTGTCGCCAAGACACAGGTTTACGATGACAA 620
Db 698 ACATGACTCCAAAGTGGTTCAAAGAGCCCTTACTTTGCTCCATGCGGTGAGTGGGGCAGCA 757
Qy 621 GATCTACTACTTTCGAGAGGACAACTCTGACAGAAATCTGAGGCTCTCTCAATGT 580
Db 758 TGCTACTTCTTTCGCGGAGTTGCGATGGAGTTTAACTACCTTGGGAGAGTGGTGT 817
Qy 681 GTCCCGTGTGGCCCACTGTGTCAGGGGGGACGAGGTTGGGAAATTCACATGTCAGTCTC 740
Db 818 GTCCCGTGTGGCCGAGTGTGCAAGACGAGTGGGAGCTCCCCCGGTGCTGGAGAA 877
Qy 741 CAAGTGAACACTTTTCTGAAAGCCATCTGCTATGTCAGTATGCTGC---CACCAACAA 797
Db 878 GCAGTGGAGCTCTCTCTGAAGCGCGCTCAACTGCTGTGTACCCGGAGACTCCCAATT 937
Qy 798 GAACCTCAACAGGCTGCAAGAGCTTCTCTGCTCCCTGACCCAGCGGCGAGTGGAGGA 857
Db 938 CTACTTCAACGTGTCAGAGGTGTCACGGGGTGTGTCAGCTCGGGGGCCGCGGTGT 997
Qy 858 CACCAAGGTCTATGTGTTTCTCAACCCCTGGAACACTCAAGCTGCTGCTGTATTC 917
Db 998 CTGCGCGTTTTCACGCCCGCAGCAACAGCATCCCTGCTGCTGCTGCTGCTTGA 1057
Qy 918 CCT 920
Db 1058 CCT 1060

RESULT 5

T72107
ID T72107 standard; cDNA to mRNA; 3692 BP.
AC T72107;
DT 13-FEB-1998 (first entry)
DE Rat semaphorin 2 gene.
KW Semaphorin 2; central nerve extension; rat; human; inhibitor;
KW central nerve regeneration promoter; ss.
OS Rattus norvegicus.
FH Key Location/Qualifiers
FT 5'UTR 1..18

FT CDS
FT 19..2682 /*tag= a
FT /*tag= b
FT 2683..3653 /*tag= c
FT /*tag= d
FT polyA_signal 3654..3692
PN W09720928-A1.
PD 12-JUN-1997.
PF 02-DEC-1996; J03517.
PF 31-OCT-1996; JP-307205.
PR 06-DEC-1995; JP-345187.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI; 97-319775/29.
DR P-ESDB; W19856.
PT Semaphorin 2 and the gene encoding it - also inhibitors of its
PT action which can be used as promoters of central nerve regeneration
PS Claim 1; Page 61-63; 106pp; Japanese.
CC This sequence represents the rat semaphorin 2 gene. The protein encoded
CC by this sequence, and the human semaphorin 2 protein (see W19857) can be
CC used for screening possible candidates for activity as semaphorin 2
CC inhibitors. Proteins containing all or part of the semaphorin 2 sequence
CC are useful as central nerve extension inhibitors. Semaphorin 2 inhibitors
CC identified by screening can be used as promoters of central nerve
CC regeneration.
SQ Sequence 3692 BP; 689 A; 1126 C; 1122 G; 755 T;

Query Match 2.5%; Score 48.4; DB 1; Length 3692;
Best Local Similarity 48.9%; Pred. No. 0.019;
Matches 130; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
Qy 522 CCGGGGCGAGAGTGAGCTGTACACAGTGATCTCATGTGTCAGAGAACACACAGTTTCATCAA 581
Db 642 CCGTAGCTTGGGACCGGCCACACTCGGCACAGTAAGCATGACTCCAGTGGTTTAA 701
Qy 582 AGCCACCATCTGTCACCAAGACAGGCTTACGATGACAGATCTACTACTTCTCCGAGA 641
Db 702 AGAGCATACTTTGTGTCATGCGGTGGAGTGGGGAAGCCACGCTACTTCTTCTCCGGGA 761
Qy 642 GGACAATCTCACAAGAACTCTGAGGCTCTCTCATGTGTCCTGTCCTGTCGCGGCTGTTG 701
Db 762 GATCCCATGAGATTAATCTGTGAAAAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCG 821
Qy 702 CAGGGGACAGGAGTGGGAAAGTTCACTGTCACTCTCCAGTGTGGAACACACTTTTCTGAA 761
Db 822 CAAGAAATGATGTGGCGGCTCCCGACCGGTCTGGAGAGCAGTGGAGTCTCTTCTCTGAA 881
Qy 762 AGCCATGCTGTGATGATGATGCTG 787
Db 882 GGCCTGGCTCAACTGCTCCGTGCTG 907

RESULT 6

Q60748
ID Q60748 standard; DNA; 265 BP.
AC Q60748;
DT 16-MAR-1994 (first entry)
DE Human brain expressed sequence tag EST00865.
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
OS Homo sapiens.
PN W09316178-A.
PD 19-AUG-1993.
PF 12-FEB-1993; U01294.
PR 12-FEB-1992; US-837195.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
DR WPI; 93-272882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes


```
AC V28915;
DE 02-SEP-1998 (first entry)
DE Human semaphorin Y encoding cDNA.
KW Human; semaphorin Y; nerve extension inhibitor; central nervous system;
KW peripheral nerve growth; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 1..187
FT 5'UTR
FT CDS
FT /*tag= a
FT 188..2980
FT /*tag= b
FT /*product= "semaphorin Y"
FT 2978..3407
FT /*tag= c
FT 3408..3432
FT /*tag= d
FT polyA_site
PN WO9811216-A1.
PD 19-MAR-1998.
PF 09-SEP-1997; J03167.
PR 08-AUG-1997; JP-227220.
PR 11-SEP-1996; JP-263565.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR P-PSDB; W57260.
DR WPI; 98-250958/22.
DR DNA encoding human and rat semaphorin Y - an inhibitor of nerve
PT extension
PS Claim 2; Page 60-62; 85pp; Japanese.
CC The present sequence encodes human semaphorin Y, which inhibits nerve
CC extension. Semaphorin Y genes and proteins may be used to inhibit
CC peripheral nerve growth. Semaphorin Y antagonists can be used to
CC accelerate regrowth of the central nervous system.
CC Sequence 3432 BP; 553 A; 1189 C; 1016 G; 674 T;
SQ Sequence 3432 BP; 553 A; 1189 C; 1016 G; 674 T;

Query Match 2.4%; Score 45.2; DB 1; Length 3432;
Best Local Similarity 49.6%; Pred. No. 0.1;
Matches 116; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 554 CTGTCATGCAGAACCCACAGTTTCATCAAGCCACCATCGTGCCACCAAGAGGCTTACG 613
DB 825 CGCCCAAGTATGACTCCCAAGTGGCTCCGAGAGCCACACTTTGTCCAGGCGCTTGGAGCATG 884
QY 614 ATGACAAGATCTACTACTTCTCCGAGGAGGACAATCCTGACAAAGATCTGTGAGGCTCCTC 673
DB 885 GAGACCATGTCCTACTTCTTCCGCGAGGTCCTGTGGAGGATCTCGGCTGGGGAGG 944
QY 674 TCAATGTGTCCCGTGTGGCCCAAGTTGTGCGAGGGGACCCAGGTCGGTGAAGATCTTCACTGT 733
DB 945 TCCAGTTCTCCGCGTAGCCCGAGTATGTAACGTGACATGGCGGCTCGCGGCT 1004
QY 734 CAGTCTCCAAGTGGAAACACTTTTCTGAAGCCATCGTGTGATGATGATGCTG 787
DB 1005 TGGACCGCCACTGGACATCTTCTTCCGAGCTTCGAGCTCAACTGCTGTGCCCTG 1058

RESULT 10
V28916
ID V28916 standard; cDNA to mRNA; 2790 BP.
AC V28916;
DT 02-SEP-1998 (first entry)
DE Human semaphorin Y encoding cDNA, CDS only.
KW Human; semaphorin Y; nerve extension inhibitor; central nervous system;
KW peripheral nerve growth; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 1..2790
FT 5'UTR
FT CDS
FT /*tag= a
FT /*product= "semaphorin Y"
FT /*note= "no stop codon given"
PN WO9811216-A1.
PD 19-MAR-1998.
PF 09-SEP-1997; J03167.

us-09-041-236-1_copy_1_1902.rng
```

```
PR 08-AUG-1997; JP-227220.
PR 11-SEP-1996; JP-263565.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI; 98-250958/22.
DR P-PSDB; W57260.
DR DNA encoding human and rat semaphorin Y - an inhibitor of nerve
PT extension
PS Claim 2; Page 63-65; 85pp; Japanese.
CC The present sequence encodes human semaphorin Y, which inhibits nerve
CC extension. Semaphorin Y genes and proteins may be used to inhibit
CC peripheral nerve growth. Semaphorin Y antagonists can be used to
CC accelerate regrowth of the central nervous system.
CC Sequence 2790 BP; 435 A; 957 C; 839 G; 559 T;
SQ Sequence 2790 BP; 435 A; 957 C; 839 G; 559 T;

Query Match 2.4%; Score 45.2; DB 1; Length 2790;
Best Local Similarity 49.6%; Pred. No. 0.097;
Matches 116; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 554 CTGTCATGCAGAACCCACAGTTTCATCAAGCCACCATCGTGCCACCAAGAGGCTTACG 613
DB 638 CGCCCAAGTATGACTCCCAAGTGGCTCCGAGAGCCACACTTTGTCCAGGCGCTTGGAGCATG 697
QY 614 ATGACAAGATCTACTACTTCTTCCGAGGAGGACAATCCTGACAAAGATCTTGTGAGGCTCCTC 673
DB 698 GAGACCATGTCCTACTTCTTCCGCGAGGTCCTGTGGAGGATCTCGGCTGGGGAGG 757
QY 674 TCAATGTGTCCCGTGTGGCCCAAGTTGTGCGAGGGGACCCAGGTCGGTGAAGATCTTCACTGT 733
DB 758 TCCAGTTCTCCGCGTAGCCCGAGTATGTAACGTGACATGGCGGCTCGCGGCT 817
QY 734 CAGTCTCCAAGTGGAAACACTTTTCTGAAGCCATCGTGTGATGATGATGCTG 787
DB 818 TGGACCGCCACTGGACATCTTCTTCCGAGCTTCGAGCTCAACTGCTGTGCCCTG 871

RESULT 11
V44295
ID V44295 standard; cDNA; 3550 BP.
AC V44295;
DT 06-OCT-1998 (first entry)
DE Human secreted protein clone CJ145_1 cDNA.
DE Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 126..3053
FT CDS
FT /*tag= a
FT /*product= secreted protein
FT /*note= "isolated from clone CJ145_1"
PN WO9827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GEM) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 98-362774/31.
DR P-PSDB; W64221.
DR New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17a; Page 69-71; 110pp; English.
CC This sequence encodes a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
```


CC Turner's syndrome. The products can be used to develop agents for the
CC treatment of short stature or other human growth disorders. The products
CC can also be used for providing a mitogenic effect on cells, e.g. for the
CC treatment of bone diseases such as osteoporosis and diseases involved
CC with disturbance in the bone calcium regulation.
SQ Sequence 32367 BP; 7627 A; 8130 C; 8564 G; 8043 T;

Query Match 2.3%; Score 43.2; DB 1; Length 32367;
Best Local Similarity 51.7%; Pred. No. 0.52;
Matches 124; Conservative 0; Mismatches 113; Indels 3; Gaps 1;
QY 1561 AAGGTTTCCTGGCCCAAACTCGCTACTACCTGAGTGGCCCAATGGAATCCGCCAC 1620
DB 13529 AAGGTACACAGTCCCAAGCCCTCTCCAGGACACACAGGTCCCAAGGCTCTCCCGGGAC 13470
QY 1621 GGCACCTACTGATGCGCCGACAGAGAACTGGAGCAGAGTGGCAACTGGTCACAC 1680
DB 13469 ACCAGATCCCAAGGCCCTCCCGGTACACAGGTGTCCAAGCTCTTCCAGAGACAC 13410
QY 1681 AGCCCCAATGATCTGTTCATCGAGAACCCTACCGGCGCAGAGTACGGCCACTACTTC 1740
DB 13409 GTCCCCAAGCCCTCTCCAGAACACACAGGTCCCAAGCCTCTCCAGAA---CACCA 13353
QY 1741 TCGAGGCCGACGAGGCGCTCTACTTCCGCGAGGCTCAGCACTGGCAGCTGTGCCG 1800
DB 13352 TCGAAGCTCTCAGGTACACAGGTCCCAAGCCCTCTCCAGGACACAGGTCCCAAG 13293

RESULT 14
ID T66541 standard; DNA; 936 BP.
AC T66541; 1997 (first entry)
DE 20K-cellulase gene from Melanocarpus albus.
KW 20K-cellulase; endoglucanase; bio-stoning; bio-washing; denim;
KW detergent; textile; pulp; paper; Myriococcus albus;
KW Thielavia albus; ss
OS Melanocarpus albus ALK04237 (CBS 685.95).
FH Key Location/Qualifiers
FT exon 33..115
FT intron 116..196
FT exon 187..435
FT intron 436..505
FT exon 506..881
FT signal_peptide 33..93
FT WO9714804-A1.
PD 24-APR-1997.
PF 17-OCT-1996; F10550.
PR 28-JUN-1996; US-020840.
PR 17-OCT-1995; US-005335.
PR 04-DEC-1995; US-007926.
PA (PRIM-) PRIMALCO LTD.
PI Elovainio M, Haakana H, Joutsjoki V, Lantto R, Londenborough J;
PI Mankylaa A, Miettinen-Oinonen A, Paloheimo M, Suominen P;
PI Vehmaapera J;
PI WPI; 97-245119/22.
DR P-PSDB; W16542.
PT DNA encoding 20K, 50K and 50K B cellulase and facilitating protein -
PT useful for bio-stoning and bio-finishing cotton fabrics, especially
PT denim, and as detergent for treating cellulosic fibre containing
PT textiles
PS Claim 1; Fig 19; 204pp; English.
CC This DNA molecule codes for a 20 kDa cellulase (W16542) with
CC endoglucanase activity (20K-cellulase) from Melanocarpus albus
CC ALK04237. The enzyme can be used over a wide range of pH and

CC temperatures up to at least 70 deg. The gene was obtained by PCR
CC amplification using primers (see T66545-46) based on tryptic
CC fragments (W16563 and W16573) of the isolated 20K-cellulase. It
CC has been deposited as ATCC 11024 and DSM 11012. It can be used to
CC prepare recombinant 20K-cellulase in transformed host cells, esp.
CC Trichoderma reesei. A novel enzyme preparation includes at least
CC 1 cellulase from Melanocarpus (see also W16543-45), Myriococcus,
CC Sporotrichum, Myceliophthora or Chaetomium. It can be used to
CC bio-stone cotton-containing fabrics or garments, especially denim,
CC or bio-finish textiles containing natural or man-made cellulose-
CC containing fibres. It can also be used in detergent, especially to
CC treat cellulosic fibre-containing textiles, to treat wood-derived
CC pulp or fibre or to improve animal feed quality.
SQ Sequence 936 BP; 147 A; 351 C; 277 G; 161 T;

Query Match 2.2%; Score 42; DB 1; Length 936;
Best Local Similarity 47.7%; Pred. No. 0.4;
Matches 123; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
QY 1608 GGAATCCCGCCAGCCACCTACTATGCGCCGACAGAACTGGAGCAGAGTGGCA 1667
DB 193 GGACTGCTGCAAGCCGCTCGTGGCGGTGGCGGAAAGGGCCCGTGAACCCGCTCTA 252
QY 1668 ACCTGTCACAGAGACCCCACTGCTGTTCATCGAGAACCCTACGCGCCACGATA 1727
DB 253 CTCGTGCGACGCGCACTTCCAGCGCATCCAGACTTCGATGCGGCTCGGAGGG 312
QY 1728 CGGCCACTACTTTCGCGAGGCGCCAGGCGCTCTTCCGCGAGGCTCAGCACTGGCA 1787
DB 313 CGGCCCGCCCTCTCGTGGCGCGACACAGCCCTTGGCCATTATGACAACTCTCGTA 372
QY 1788 GCTGTGCGCGAGGACGATCATGCGCGAGCACCTGTGTGGTTCATGCTGTGCCCTG 1847
DB 373 CGGCTTCGCGGCGACTGCACCTCAGCGCCGAGAGTGTGTGTGCTGTGCTGCTA 432
QY 1848 CGCCTCCCTCTGCTGGTGGG 1865
DB 433 CGCGTGTGCTGTGGG 450

RESULT 15
ID T93095 standard; cDNA; 24379 BP.
AC T93095;
DE Streptomyces frenolicin gene cluster.
KW Frenolicin; antibiotic; feed additive; anticoccidial;
KW coccidiostatic; efflux pump; butyrate starter synthase;
KW polyketide synthase; PKS; hemiketalase; ketoreductase; cyclase;
KW dehydrase; ketoreductase; hydroxylase; Streptomyces roseofulvus;
OS Streptomyces sp.
FH Key Location/Qualifiers
FT CDS 636..2948
FT /*tag= a
FT /*product= 80 kDa non-membrane protein
FT /*note= "gene A (specifically claimed)"
FT CDS 2945..3916
FT /*tag= b
FT /*product= membrane protein
FT /*note= "gene B (specifically claimed)"
FT CDS 4020..4844
FT /*tag= c
FT /*product= protein with 6 membrane-spanning domains
FT /*note= "gene C (specifically claimed)"
FT CDS 4841..6415
FT /*tag= d
FT /*product= ATP-binding component of ABC transporter
FT /*note= "gene D (specifically claimed)"
FT CDS 6533..7183
FT /*tag= e
FT /*product= unknown non-membrane protein

FT	/note= "gene E (specifically claimed)"
CDS	7344. .8897
FT	/*tag= f
FT	/product= putative efflux pump
FT	/note= "gene F (specifically claimed)"
FT	9164. .10012
FT	/*tag= g
FT	/product= transcription activator
FT	/note= "gene G (specifically claimed)"
CDS	complement (10105. .10621)
FT	/*tag= h
FT	/product= translationally coupled to gene I
FT	/note= "gene H (specifically claimed)"
CDS	complement (10618. .11628)
FT	/*tag= i
FT	/product= homologue of fabH
FT	/note= "gene I (specifically claimed)"
FT	11809. .12066
FT	/*tag= j
FT	/product= acyl carrier protein
FT	/note= "gene J (specifically claimed)"
CDS	complement (12154. .13209)
FT	/*tag= k
FT	/product= acyltransferase
FT	/note= "gene K (specifically claimed)"
CDS	13409. .14686
FT	/*tag= l
FT	/product= PKS ketoacylsynthase subunit
FT	/note= "gene L (specifically claimed)"
FT	14767. .16047
FT	/*tag= m
FT	/product= PKS chain length factor
FT	/note= "gene M (specifically claimed)"
CDS	16120. .16371
FT	/*tag= n
FT	/product= acyl carrier protein used by the PKS
FT	/note= "gene N (specifically claimed)"
CDS	complement (16453. .16935)
FT	/*tag= o
FT	/product= putative hemiketal dehydrase
FT	/note= "gene O (specifically claimed)"
FT	17088. .17903
CDS	/*tag= p
FT	/product= ketoreductase related to actIII
FT	/note= "gene P (specifically claimed)"
CDS	17903. .18898
FT	/*tag= q
FT	/product= cyclase/dehydrase related to act VII
FT	/note= "gene Q (specifically claimed)"
FT	18895. .19839
CDS	/*tag= r
FT	/product= cyclase/dehydrase related to actIV
FT	/note= "gene R (specifically claimed)"
CDS	complement (19990. .20907)
FT	/*tag= s
FT	/product= oxidoreductase
FT	/note= "gene S (specifically claimed)"
CDS	complement (20904. .22094)
FT	/*tag= t
FT	/product= quinone-forming hydroxylase
FT	/note= "gene T (specifically claimed)"
FT	complement (22505. .22179)
FT	/*tag= u
FT	/product= not specified
FT	/note= "gene U"

EP-806480-A2.

PN EF-800480-AZ
PD 12-NOV-1997.

PF 02-MAY-1997; 107329.

PR 07-MAY-1996; US-016753.

PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
07 MAR 1996, 03 010735.

PI Reeves CD, Soliday CL;

DR WPI; 97-538619/50.

DR P-PSDB; W34199-W34219.

PT	Streptomyces frenolicin gene cluster - useful for producing
PR	recombinant frenolicin antibiotics
PS	Claim 1; Page 40-60; 66pp; English.
CC	This DNA sequence comprises the Streptomyces frenolicin gene
CC	cluster containing specifically claimed coding sequences (genes
CC	A-U) that respectively encode 21 proteins (see W34199-219) involved
CC	in frenolicin synthesis. The genes can be divided into 5
CC	subclusters: (1) genes A, B, C, D and F encode an efflux pump; (2)
CC	genes H, I, J and K encode butyrate starter synthases; (3) genes
CC	L, M and N encode polyketide synthases (PKS); (4) genes O, P, Q and
CC	R encode a hemiketalase, a ketoreductase and cyclases/dehydrases;
CC	and (5) genes S and T encode a keto/enoyl reductase and a
CC	hydrolase. Also claimed are vectors, host cells (especially a
CC	Streptomyces sp., particularly Streptomyces roseofulvus), and the
CC	encoded proteins. Cells transformed using the above sequence can
CC	be cultured to produce frenolicins or frenolicin precursors. The
CC	precursors can be converted to frenolicins by chemical or other
CC	methods. The frenolicins can be oxidised to frenolicin B, an
CC	antibiotic used as an anticoccidial agent. The frenolicins can be
CC	used as animal feed additives.
SQ	Sequence 24379 Bp; 3077 A; 9792 C; 8499 G; 3011 T;
Query Match 2.2%; Score 42; DB 1; Length 24379;	
Best Local Similarity 48.3%; Pred. No. 0.91;	
Matches 117; Conservative 0; Mismatches 125; Indels 0; Gaps	
QY	1534 CCACAAACCAGACAAGCCCCACTGCAGAGAAGTTTCCTCGGCCCAAACTCTCGTACTTAC 1593
DB	2517 CTGAGCCCGCTCTTCCTCGGGCAGCAGAGCCTCGAATCTGTGCAGCACAGCTCAAGAC 2576
QY	1594 CTGAGTGTCGCCCATGGAATCCGCCACGCCACCTACTCATATGGCCACACAGGAGAAGCTG 1653
DB	2577 ATCGGGCTCGAACTGCGGCTGAGACAGCTCACCGTCTCCCCTGTTCCCGGAAGTGCTCGCC 2636
QY	1654 GAGCAGAGCTCGAACCTGGTCACAGAGCCCACTGCATCTCTGTTTCATCGAGAAGCTC 1713
DB	2637 GCGGGCAGCTTACGACCTGAGCCTCCAGAGCGGCACACCGCGCGATCCCGAGCTCCTCAC 2696
QY	1714 ACGGCGCAGCAGTAGCGGCCACTACTTCTCGAGGGCCACGAGGAGGCTCTCTACTTCCGGGAG 1773
DB	2697 ACGGCTTCGCGGGCGCGCACCCCGCTGCGCCAGCGCGGCTTCGCGTCCGAAGCTCGCGGG 2756
QY	1774 GC 1775
DB	2757 GC 2758

Search completed: July 8, 1999, 19:18:57
Job time: 2126 sec

Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 1737 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 431.

FEATURES

Source
1..476
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand
cDNA was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCCGCTCATTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
BASE COUNT 106 a 134 c 147 g 89 t
ORIGIN

Query Match 23.8%; Score 452.4; DB 27; Length 476;

Best Local Similarity 99.4%; Pred. No. 1.1e-109; Mismatches 1; Indels 2; Gaps 2;

Matches 475; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 141 CCACGAGCCAGGAGCTCTCTGTGGTGGGAGGACGTGGCAAGTCTACCTCTTTGA 200
|||||
Db 1 CCAGAGCCAGGAGCTCTCTGTGGTGGGAGGACGTGGCAAGTCTACCTCTTTGA 60
|||||
QY 201 CTTCCTGGAGGCAAGAACGCATCTGTGGCAGCGTGAATATCGGTTCCACAAAGGGTC 260
|||||
Db 61 CTTCCTGGAGGCAAGAACGCATCTGTGGCAGCGTGAATATCGGTTCCACAAAGGGTC 120
|||||
QY 261 CTCTCTGGATAAGCGGAGCTCCGAGAACTACATCACTCTCTGGAGGCGGAGTGAGG 320
|||||
Db 121 CTCTCTGGATAAGCGGAGCTCCGAGAACTACATCACTCTCTGGAGGCGGAGTGAGG 180
|||||
QY 321 GCTGCTGGCTGTGGACCAACCCCGGACCCAGCTGTGGAACTGTGTAATGGCAC 380
|||||
Db 181 GCTGCTGGCTGTGGACCAACCCCGGACCCAGCTGTGTAATGGCAC 240
|||||
QY 381 TGTGGTGCCACTTGGGAGATGAGAGCTACGCCCTTCAGCCCGGACGAGAACTCCCT 440
|||||
Db 241 TGTGGTG-CACTTGGGAGATGAGAGCTACG-CCCTTCAGCCCGGACGAGAACTCCCT 298
|||||
QY 441 GGTCTCTTTGAAGGGGACGAGTGTATCCACCATCCGGAACGAGTAACATGGAA 500
|||||
Db 299 GGTCTCTTTGAAGGGGACGAGTGTATCCACCATCCGGAACGAGTAACATGGAA 358
|||||
QY 501 GATCCCTCGGTTCCCGGCAATCCGGGCGAGAGTGAGCTGTACACAGTGATAGTCTCAT 560
|||||
Db 359 GATCCCTCGGTTCCCGGCAATCCGGGCGAGAGTGAGCTGTACACAGTGATAGTCTCAT 418
|||||
QY 561 GCAGAACCCACAGTTTCATCAAGCCACCATCGTGCCACCAAGACGAGCTTACGATGAC 618
|||||
Db 419 GCAGAACCCACAGTTTCATCAAGCCACCATCGTGCCACCAAGACGAGCTTACGATGAC 476
|||||

RESULT 2

FEATURES

Source
1..459
/organism="Homo sapiens"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'-
TGTTACCAATCTGAAGTGGAGCGCCGCTCATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="Soares_placenta_8to9weeks_2NBHP8to9"
/dev_stage="two placenta: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT 87 a 134 c 151 g 87 t
ORIGIN

Query Match 23.6%; Score 449.4; DB 23; Length 459;

Best Local Similarity 98.7%; Pred. No. 6.8e-109; Mismatches 6; Indels 0; Gaps 0;

Matches 453; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 19 GCGGCGCGGCTCCGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 78
|||||
Db 1 GCGGCGCGGCTCCGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 60
|||||

QY 79 AAAGGCCATGTAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 138
|||||
Db 61 AAAGGCCATGTAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
|||||

QY 139 TTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 198
|||||
Db 121 TTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
|||||

QY 199 GACTTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 258
|||||
Db 181 GACTTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
|||||

QY 259 TCCTGCTGTGATGAAGGGGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 318
|||||
Db 241 TCCTGCTGTGATGAAGGGGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
|||||

QY 319 GGGCTGTGGCTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 378
|||||

```
|||||
Db 301 GGGTCTGGCTGGCCACCAAGCCGCCAGCTGCTGGAACCTGCTGAATGCG 360
QY 379 ACTGTGGTCCCACTTGGCCAGATGAGAGCTAGGCCCTTACGCCCGGACGAGAACTCC 438
Db 361 ACTGTGGTCCCACTTGGCCAGATGAGAGCTATGCCCCCTTCAGTCCGGACGAGAACTAC 420
QY 439 CTGGTCTCTTTGAAGGGGACGAGGTGATTCCACCATC 477
Db 421 CGTGTCTGTTGAAGGGGACGAGGTGATTCCACCATC 459

RESULT 3
LOCUS AI095332 469 bp mRNA EST 24-SEP-1998
DEFINITION oyl7c05.s1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA
clone IMAGE:1666088 3' similar to TR:Q64906 Q64906 SIMILAR TO
GENBANK ACCESSION NUMBER L26081. ;, mRNA sequence.
ACCESSION AI095332
NID 93434308
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 469)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1645 Std Error: 0.00
Seq primer: 40ml3 fwd. ET from Amersham
High quality sequence stop: 468.
Location/Qualifiers
1. 469
/organism="Homo sapiens"
/clone="vector: pT73D (Pharmacia) with a modified
polylinker V_type: phagemid; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTT 3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1666088"
/clone_lib="Soares_senescent_fibroblasts_NbHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT 91 a 136 c 156 g 86 t
ORIGIN

Query Match 21.8%; Score 413.8; DB 30; Length 469;
Best Local Similarity 97.7%; Pred. No. 1.8e-99;
Matches 462; Conservative 0; Mismatches 7; Indels 4; Gaps 4;

QY 19 GCGCGCGCGCCCTCGCCCGACCTTGAAGAGCGGACCCCGCATCTTCCCGCTCTGG 78
Db 1 GCGCGCGCGCCCTCGCCCGACCTTGAAGAGCGGACCCCGCATCTTCCCGCTCTGG 60
QY 79 AAAGGCCATGTAGGCGAGGCGGTGGACTTTGGCCAGACTGAGCGGACGAGTGT 138
Db 61 AAAGGCCATGTAGGCGAGGCGGTGGACTTTGGCCAGACTGAGCGGACGAGTGT 120
```

```
QY 139 TTCCACGAGCAGGAGCTCTCTCTGTGGTGGGAGACGTGGCAAGGTCTACCTCTTT 198
Db 121 TTCCACGAGCAGGAGCTCTCTCTGTGGTGGGAGACGTGGCAAGGTCTACCTCTTT 180
QY 199 GACTTCCCGGAGGGAAGAAGCATCTCTGTGGCAGCGTGAATATCGGTCCACAAAGGG 258
Db 181 GACTTCCCGGAGGGAAGAAGCATCTCTGTGGCAGCGTGAATATCGGTCCACAAAGGG 240
QY 259 TCCTGTCTGGATAAGCGGGACTTGGCAGAACTACATCACTCTCTCTGGAGAGCGGAGTGAG 318
Db 241 TCCTGTCTGGATAAGCGGGACTTGGCAGAACTACATCACTCTCTCTGGAGAGCGGAGTGAG 300
QY 319 GGGTCTGTGGCTGTGGCAGCAACCCCGGACCCACCTGCTGGAACCTGCTGAATGCG 378
Db 301 GGGTCTGTGGCTGTGG-ACCAACCGCGGACCCACCTGCTGGAACCTGCTGAATGCG 359
QY 379 ACTGTGGTGGCACTTGGCGAGATGAGAGCTACGCCCTTTCAGCCCGGAGAGACTCC 438
Db 360 ACTGTGGT-CATCTTGGCGAGATGAGAGCTACG-CCCTTTCAGCCCGGAGAGACTCC 417
QY 439 CTGTTCTCTTTGAAGGGGAGGAGGTGATTCCACCATCCGGAAGCAGGAATA 491
Db 418 CTGTTCTCTTTGAAGGGGAGGAGGTGATTCC-CCATCCGGAAGCAGGAATA 469

RESULT 4
LOCUS W47265/c 424 bp mRNA EST 11-OCT-1996
DEFINITION ze39a12.r1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA
clone 324670 5', mRNA sequence.
ACCESSION W47265
NID g1331973
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; mitochondria eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 424)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 731 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 394.
Location/Qualifiers
1. 424
/organism="Homo sapiens"
/clone="vector: pT73D (Pharmacia) with a modified
polylinker V_type: phagemid; Site_1: Not I; Site_2: Eco
RI; TGTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTT 3',
double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT73
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="324670"
```

FEATURES

source

mRNA	/clone_lib="Soares senescent fibroblasts MbHSP"			
	/tissue_type="senescent fibroblast"			
	/lab_host="DH10B (ampicillin resistant)"			
	<1. >424			
BASE COUNT	79 a	129 c	117 g	97 t
ORIGIN	2 others			

Query Match	19.7%	Score 374.8;	DB 14;	Length 424;
Best Local Similarity	97.2%	Pred. No. 3.5e-89;		
Matches 413; Conservative	0;	Mismatches 8;	Indels 4;	Gaps 3;

Qy	307	AGCGGAGT	GAGGGCTGCT	TGGCCCTGTGGCACC	AAGCCCGGCACCC	CCAGCTGCTGCTGGAAC	366
Db	424	AGCGGAGT	GAGGGCTGCT	-GCTGTGGCACC	AAGCCCGGCACCC	CCAGCTGCTGCTGGAAC	366
Qy	367	CTGGTGAAT	GGCAGTCTGGT	TGGCCACTTGGG	GAGATGAGAGGCT	TACGCCCTTTCAGCCCG	426
Db	365	CTGGTGANAT	GCAGTCTGGT	GTCACACTTGGC	CAGA-GTGAGG	AAGACGCCCTTTCAGCCCG	307
Qy	427	GACGAGAAC	--TCCCTGGT	TCTGTTTGAAGGG	GACGAGTGTATTCC	ACCACATCCGGAAGC	484
Db	306	GACGAGAAC	GTCCCGTGTCTG	TTGAAGGG	GACGAGTGTATTC	ACCATCCCGGAAGC	247
Qy	485	AGGAATACAT	GGGAGATCCCT	TCGGTTCGCGCG	CAATCCGGGCGG	AGAGTGTAGCTGTACA	544
Db	246	AGGAATACAT	GGGAGATCCCT	CGGTTCGCGCG	CAATCCGGGCGG	AGAGTGTAGCTGTACA	187
Qy	545	CCAGTGAT	ACTGTGCATG	CACAAACCAC	AGTTTCATCAAAG	CCACCATCTGTGCACCAAGACC	604
Db	186	CCAGTGAT	ACTGTGCATG	CACAAACCAC	AGTTTCATCAAAG	CCACCATCTGTGCACCAAGACC	127
Qy	605	AGGCTTAC	GATGACAGAT	CTACTTCTTCG	GAGGAGACAATCTC	TGCACAGAATCTCTG	664
Db	126	AGGCTTAC	GATGACAGAT	CTACTTCTTCG	GAGGAGACAATCTC	TGCACAGAATCTCTG	67
Qy	665	AGGCTCTC	TCAATGTGTCC	CGTGTGGCC	CAAGTTGTGC	AGGGGGGACACGAGGTGGGGAAA	724
Db	66	AGGCTCTC	TCAATGTGTCC	CGTGTGGCC	CAAGTTGTGC	AGGGGGGACACGAGGTGGGGAAA	7

RESULT	5
H02902	
LOCUS	446 bp mRNA
DEFINITION	H02902
ACCESSION	YJ39F01.s1 Homo sapiens cDNA clone 151129 3'
NID	9865835
KEYWORDS	EST.
SOURCE	human clone-151129 library=Soares placenta Nb2HP vector=pt7T3D

(Pharmacia) with a modified polylinker, *hOst-DNA* (ampicillin resistant) primer-Promega -21m13 Rsite1-Not I Rsite2-Eco RI Female placenta obtained at birth (full term). 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTCGAGAATTCGGCGCAGGAATTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pR73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo.

ORGANISM

constructed by *benlo* Soares and *M.rattina* bonatuo. .

Homo sapiens

Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcoterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoideae; Homo. 1 (bases 1 to 446)

REFERENCE

AUTHORS

Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasaki, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE	JOURNAL	COMMENT
The Washu-Merck EST Project	Unpublished (1995)	

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 196
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

location/individuals	source
1. .446	
/organism="Homo sapiens"	
/clone="151129"	
93 a 115 c 136 g	93 t
9 others	
BASE COUNT	
ORIGIN	

Query Match	18.4%	Score 350;	DB 11;	Length 446;
Best Local Similarity	92.8%	Pred. No. 1.3e-82;		
Matches 384;	Conservative	0;	Mismatches 28;	Indels 2; Gaps 2;

Qy	70	GC	CTCTGGAAGCCATGTAGG	CAGGACGGGTGAC	TTTGGCCAGACTGTAGCCGCAC	129
Db	1	GC	CTCTGGAAGCCATGTAGG	CAGGACGGGTGAC	TTTGGCCAGACTGTAGCCGCAC	60
Qy	130	AC	GGTGCTTTTCACGAGCCAG	CAGCAGCTCTCTGTGTGG	TGGGAGAGACGTGGCAAGGTC	189
Db	61	AC	GGTGCTTTTCACGAGCCAG	CAGCAGCTCTCTGTGTGG	TGGGAGAGACGTGGCAAGGTC	120
Qy	190	TAC	CTCTTTGACTTCCCGAGG	CGCAAGCATCTGTGCGC	ACGGTGAATATCGGCTCC	249
Db	121	TAC	CTCTTTGACTTCCCGAGG	CGCAAGCATCTGTGCGC	ACGGTGAATATCGGCTCC	180
Qy	250	ACAA	GGGGTCTGTCTGTGATAG	CGGGACTGCGAGAACTAC	ATCTCTCTGGAGAGG	309
Db	181	ACAA	GGGGTCTGTCTGTGATAG	CGGGACTGCGAGAACTAC	ATCTCTCTGGAGAGG	240
Qy	310	CGG	AGTAGGGGCTGTGGCCTGT	GGCACCAAGCCCGCACCC	ACAGCTGCTGGAACCTG	369
Db	241	CGG	ANTGAGGGGTGTGGCCTGT	GG-ACCAAGCCCGCACCC	ACAGCTGCTGGAACCTG	299
Qy	370	GT	GAATGGCACTGTGTGGC	ACATTTGGCGAGATGAG	GGCTACGCCCCCTTCAGG	CCCGAC 429
Db	300	GGT	CAATNGCACINTGGT	CACCTTGGCGAGATGAG	AGGCTACGCCCTTTTCAG	CCCGAC 359
Qy	430	GAGA	CTCCCTGGTTCTG-TTT	GAAAGGGAGAGAGGTGT	ATTCACCACTATCCGGAA	482
Db	360	GAGA	CTCTCTGGTTCTNTTTT	TAAGGGGACGAAGTTATT	TTTTCANCATCNGAA	413

RESULT	6
AA968218	
LOCUS	
DEFINITION	AA968218 422 bp mRNA EST 19-MAY-1998
	uhl4509.r1 Soares mouse hypothalamus NMH1 Mus musculus cDNA clone
	1617929 5', similar to TR:Q64906 Q64906 SIMILAR TO GENBANK ACCESSION
	NUMBER L26081. ;, mRNA sequence.

ACCESSION	AA960216
NID	93142111
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 423)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,R., Wylie,T., Lennon,G., Soares,R., Wilson,R. and others

D_b 61 GGCAGCTCCTCTGTGGGTGGGAGGACGTGGCAAGGTCTACTCTTTGACTTCCCCGAG 120

QY	211	GGCAAGAACCATCTGTGCCACGGTGAATATATCGGTCCACAAAGGGGTCTCTCTGGAT	270
Db	121	GGCAAGAACCATCTGTGCCACGGTGAATATCGGTCCACAAAGGGGTCTCTCTGGAT	180
QY	271	AACGGGACTGCCAGAACTACATCACTCTCTGAGAGGGGAGTGGGGCTGGCC	330
Db	181	AACGGGACTGCCAGAACTACATCACTCTCTGAGAGGGGAGTGGGGCTGGCC	240
QY	331	TGTGGACCAACCCCGGACCCAGCTGCTGGAACCTGGTGAATGGCACTGTGTGCCA	390
Db	241	TGTGGACCAACCCCGGACCCAGCTGCTGGAACCTGGTGAATGGCACTGTGTGCCA	300
QY	391	CTTGGGAGATGAGAGCTACGCCCTTCAGCCCG	426
Db	301	CTTGGGAGATGAGAGCTACGCCCTTCAGCCCG	336
RESULT	8		
H03806/c			
LOCUS	H03806	434 bp mRNA	EST
DEFINITION	yj39f01.r1 Homo sapiens cDNA clone 151129 5' similar to SP:A49069		20-JUN-1995
ACCESSION	H03806		
NID	9868739		
KEYWORDS	EST.		
SOURCE	human clone=151129 library=Soares placenta Nb2HP vector=pf7t3d (Pharmacia) with a modified polylinker host=DHI0B (ampicillin resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Female placenta obtained at birth (full term). 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAGAAATCGCGCCGACAGATTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 434)		
AUTHORS	Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Hollman,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.		
TITLE	The WashU-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 292 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Possible reversed clone: similarity on wrong strand. Location/Qualifiers 1. .434 /organism="Homo sapiens" /clone="151129"		
FEATURES	source		
BASE COUNT	77 a	126 c	128 g 96 t
ORIGIN	7 others		
Query Match	17.2%	Score 327.4;	DB 11; Length 434;
Best Local Similarity	91.5%;	Pred. No. 1.2e-76;	
Matches 388;	Conservative 0;	Mismatches 28;	Indels 8; Gaps 4;

QY	1024	CCCACAGAGACCTTCCAGTGGCTGACCCGTCA---CCCAGAGGTGGC--GCAGAGGGTGG	1078
Db	423	CACAAAAACCTTCCCANGTGGCTGACCCGTTNAACCCCAAGAGTGGCCCAAGAGGGTGG	364
QY	1079	AGCCCATGGGGCTCTGAA--GACGCCATTGTCCACTCTAAATACCACCTACCAGAACT	1136
Db	363	AANCCATGGGGCTCTGAAAGACGCCCATTTGTCCACTCTAAATACCACCTACCAGAACT	304
QY	1137	GGCGGTCCACCCATGCAAGCCAGCCAGCGGGGAGACCTTTTCATGTGCTTTACTTACTAC	1196
Db	303	GGCGGTTCACCCATGCAAGCCAGCCAGCGGGGAGACCTTTTCATGTGCTTTACTTACTAC	244
QY	1197	AGACAGGGGCACTATCCACAAAGGTGGTGAACCGGGGGAGCAGAGCAGCTTCGCCCTT	1256
Db	243	AGACAGGGGCACTATCCACAAAGGTGGTGAACCGGGGGAGCAGAGCAGCTTCGCCCTT	184
QY	1257	CAACATCATGGAGATCCAGCCCTTCGGCCGGCGGCTGCCATCCAGACCATGTCGCTGA	1316
Db	183	CAACATCATGGAGATCCAGCCCTTCG--CGCGCGGCTGCCATCCAGACCATGTCGCTGA	125
QY	1317	TGCTGAGCGGAGGAAGCTGTATGTGAGCTCCAGTGGGAGGTGAGCCAGGTGCCCTTGA	1376
Db	124	TGCTGAGCGGAGGAAGCTGTATGTGAGCTCCAGTGGGAGGTGAGCCAGGTGCCCTTGA	65
QY	1377	CCTGTGTGAGGTCTATGCGGGGGGCTGCCACGGTTGCCCTCATGTCCCGAGACCCCTACTG	1436
Db	64	CCTGTGTGAGGTCTATGCGGGGGGCTGCCACGGTTGCCCTCATGTCCCGAGACCCCTACTG	5
QY	1437	CGGC 1440	
Db	4	NGGC 1	
RESULT	9		
LOCUS	AA856724	332 bp mRNA	EST
DEFINITION	od34c01.s1 NCI-CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1371744 3' similar to TR:Q64906 Q64906 SIMILAR TO GENBANK ACCESSION NUMBER L26081. ; mRNA sequence.		09-MAR-1998
ACCESSION	AA856724		
NID	92945026		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 332)		
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html		
FEATURES	Seq primer: -40ml3 fwd. ET from Amersham High quality sequence stop: 323. Location/Qualifiers 1. .332 /organism="Homo sapiens" /note="vector: pf7t3d-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA source		

was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-GGTACCAATGAAGTGGGCGCGCTCAATTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pMT3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

/db_xref="taxon:9606"
/clone="IMAGE:137174"
/clone_lib="NCI.CGAP.GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH108"

BASE COUNT 66 a 95 c 109 g 62 t

Query Match 16.2%; Score 308.4; DB 27; Length 332;
Best Local Similarity 99.1%; Pred. No. 1.2e-71;

Matches 331; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 105 GGACTTTGGCCAGACTGAGCGGCACAGCGTGTCTTTCCACGAGCCAGGCGAGCTCCTCTGT 164
|||||
Db 1 GGACTTTGGCCAGACTGAGCGGCACAGCGTGTCTTTCCACGAGCCAGGCGAGCTCCTCTGT 60
|||||
QY 165 GTGGGTGGGAGGAGCTGGCAAGTCTACCTCTTTTGACTTCCCGAGGGCAAGAACGCATC 224
|||||
Db 61 GTGGGTGGGAGGAGCTGGCAAGTCTACCTCTTTTGACTTCCCGAGGGCAAGAACGCATC 120
|||||
QY 225 TGTGGCCAGCGTGAATATCGGCTCCACAAAGGGTCTGTCTGGATGAAGCGGAGCTCGCGA 284
|||||
Db 121 TGTGGCCAGCGTGAATATCGGCTCCACAAAGGGTCTGTCTGGATGAAGCGGAGCTCGCGA 180
|||||
QY 285 GAACTACATCAGCTCTCTGGAGAGGGGAGTGGGGCTCTGGCTGTGGCACCACACGC 344
|||||
Db 181 GAACTACATCAGCTCTCTGGAGAGGGGAGTGGGGCTCTGGCTGTGGCACCACACGC 240
|||||
QY 345 CCGGCACCCAGCTGTGGAACCTGGTGAATGGCACTGTGGTGGCCACTTGGCGAGATGAG 404
|||||
Db 241 ACGGCACCCAGCTGTGGAACCTGGTGAATGGCACTGTGGTG-CACCTGGCGAGATGAG 299
|||||
QY 405 AGCTAGGCCCTTACGCCCGGACGAGAACTCC 438
|||||
Db 300 AGGCTAGC-CCCTTACGCCCGGACGAGAACTCC 332
|||||

RESULT 10

R3337 344 bp mRNA EST 28-APR-1995
LOCUS yh79d03.r1 Homo sapiens CDNA clone 135941 5' similar to
DEFINITION Sp:VA39_VACCC P21062 PROTEIN ;
ACCESSION R3337
NID 9789395
KEYWORDS EST.
SOURCE human clone=135941 library=Soares placenta Nb2HP vector=pf7T3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=M13Rp1 Rsite1-Not I Rsite2-Eco RI Female placenta obtained at birth (full term). 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'.

ACCTGGAAGAATTCGGCCGAGCAATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pMT3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo.

ORGANISM

Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 344)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE

JOURNAL

COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

High quality sequence stops: 275

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Possible reversed clone: similarity on wrong strand.

FEATURES

Source

1..344

/organism="Homo sapiens"

/clone="135941"

BASE COUNT 74 a 87 c 106 g 75 t 2 others

ORIGIN

Query Match 15.4%; Score 292.4; DB 10; Length 344;

Best Local Similarity 98.0%; Pred. No. 2e-67;

Matches 337; Conservative 0; Mismatches 3; Indels 4; Gaps 4;

QY 679 GTCTCCCGTGTGGCCAGTGTTC-AGGGGGACACAGGTTGGGGAAGTTCAGTGTCACT 737
|||||

Db 344 GTCTCCCGTGTGGCCAGTGTTC-AGGGGGACACAGGTTGGGGAAGTTCAGTGTCACT 285
|||||

QY 738 CTC-CAAGTGGGAACACTTTTCTG-AAAGGCATGCTGGT-ATGCAGTGTATGCCACCAA 794
|||||

Db 284 CTC-CAAGTGGGAACACTTTTCTG-AAAGGCATGCTGGT-ATGCAGTGTATGCCACCAA 225
|||||

QY 795 CAAGAATTCAACAGGCTGCAAGAGCTTCTCTGCTCCCTGACCCCGGCGGAGTGAG 854
|||||

Db 224 CAAGAATTCAACAGGCTGCAAGAGCTTCTCTGCTCCCTGACCCCGGCGGAGTGAG 165
|||||

QY 855 GGACACAGGCTGTATGGTGTTCCTCCAAACCCCTGGAAGTGTCTGTGTGTA 914
|||||

Db 164 GGACACAGGCTGTATGGTGTTCCTCCAAACCCCTGGAAGTGTCTGTGTGTA 105
|||||

QY 915 TTCCCTCGGTGACATTGACAAGGTCTTCGTAAGTCTCTCACTCAAGGGGTACCACTCAAG 974
|||||

Db 104 TTCCCTCGGTGACATTGACAAGGTCTTCGTAAGTCTCTCACTCAAGGGGTACCACTCAAG 45
|||||

QY 975 CCTTCCCAACCGCGCGCTGGCAAGTGTCTCCAGACACGAGCAGC 1018
|||||

Db 44 CCTTCCCAACCGCGCGCTGGCAAGTGTCTCCAGACACGAGCAGC 1
|||||

RESULT 11

AA260340

LOCUS

DEFINITION

AA260340 347 bp mRNA EST 18-MAR-1997

vag1b02.r1 Soares mouse 3NME12 5 Mus musculus CDNA clone 746763 5'

similar to TR:G100717 G100717 SIMILAR TO GENBANK ACCESSION NUMBER

L26081.; mRNA sequence.

ACCESSION

AA260340

NID

91896824

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Mus.

REFERENCE

1 (bases 1 to 347)

AUTHORS

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, N., Tan, F., Underwood, K., Moore, B.,

Thiesing, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:455747
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 296.
Location/Qualifiers

FEATURES

source
1. .347
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: p7T73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCCTATTTTTTTTTTTT 3'],
on total mouse RNA (provided by Minoru Ko, Wayne State
Univ.); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p7T73 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaudo."
/db_xref="taxon:10090"
/clone="746763"
/clone_lib="Soares mouse 3NME12 5"
/sex="unknown"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
<1. .347
75 a 102 c 90 g 80 t

mRNA
BASE COUNT
ORIGIN

Query Match 14.9%; Score 283; DB 16; Length 347;
Best Local Similarity 88.5%; Pred. No. 6.2e-65;
Matches 307; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 638 GAGAGACAATCCTGACAGAATCTGAGGCTCTCTCAATGTGTCGCGTGGCCAGT 597
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GAGAAGACAACCTGACAGAACCCCGAGGCTCTCTCAATGTGTCGCGAGTACCCAGT 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 698 TGTGAGGGGGGACAGGGTGGGAAAGTTCACTCTGAGTCTCCAAAGTGGACACTTTC 757
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 TGTGAGGGGGGACAGGGTGGTGAAGTTCGTTGTCTCTCAAGTGGACACTTCC 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 758 TGAAGACCATGCTGGTATGACGTGTCGTCACCAACCAAGAACTTCAACAGGCTGCAAG 817
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 TGAAGACCATGTTGCTGTCAGCGATGACGACCAACCAAGAACTTCAATCGCTGCAAG 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 818 AGCTTCTCTGCTGCTGACCCAGGCGGACGTGAGGGACACAGGCTCTATGCTGTTT 877
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 ATGTCTTCTGCTGCTGACCCAGTGGCGAGTGGAGAGATACCAAGGCTCTATGCGGTTT 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 878 TCTCAACCCCTGGAACTACTCAGCGTGTGTGTTATTCCTCGGTGACATTGACAAGG 937
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TCTCAACCCCTGGAACTACTCAGCTGTCTGCGGTGATTTCGTTGGTGACATTGACAGAG 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 938 TCTTCGCTACCTCTCAAGGGCTACCACTCAAGGCTTCCCAAC 984
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TCTTCGCTACCTCATCGCTCAAGAGCTACCAATGGGCTTTCACC 347
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12

R33439 309 bp mRNA EST 28-APR-1995
LOCUS YH79d03.s1 Homo sapiens cDNA clone 135941 3'.
DEFINITION R33439
ACCESSION R33439
NID 6789297
KEYWORDS EST.
SOURCE human clone=135941 library=Soares placenta NB2HP vector=p7T73D
(Pharmacia) with a modified polylinker host=DH10B (ampicillin
resistant) primer=Promega -2ml3 Rsite1-Not I Rsite2-Eco RI Female
placenta obtained at birth (full term). 1st strand cDNA was primed
with a Not I - oligo(dT) primer [5'
ACTGGAAGAATCGGCGCGAGAGAAATTTTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the modified p7T73
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaudo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 309)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfs, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 273
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyT not found.

FEATURES

source
1. .309
/organism="Homo sapiens"
/clone="135941"
BASE COUNT 58 a 86 c 104 g 58 t 3 others
ORIGIN

Query Match 14.8%; Score 282.2; DB 10; Length 309;
Best Local Similarity 97.7%; Pred. No. 9.7e-65;
Matches 295; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 32 CGGCCAGGGCCACCTAAGGAGCGGACCCGCATCTTCGCGCTGGAAAGCCATGTAG 91
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8 CGGCCAGGGCCACCTAAGGAGCGGATNCTANNTCTTCGCGCTGGAAAGCCATGTAG 67
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 92 GCGAGGACCGGGTGACTTTGGCCAGACTGAGCCGACACGCTGCTTTCCACGAGCCAG 151
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68 GCGAGGACCGGGTGACTTTGGCCAGACTGAGCCGACACGCTGCTTTCCACGAGCCAG 127
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 152 GCAGCTCCTCTGTGTGGGTGGGAGGAGTGGCAAGGTCTACCTCTTTGACTTCCCGGAGG 211
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 128 GCAGCTCCTCTGTGTGGGTGGGAGGAGTGGCAAGGTCTACCTCTTTGACTTCCCGGAGG 187
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 212 GCAAGACCGCATCTGTGGCCAGCGTGAATATCGGCTCCCAAGGGGTCTCTGTGATTA 271
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188 GCAAGACCGCATCTGTGGCCAGCGTGAATATCGGCTCCCAAGGGGTCTCTGTGATTA 247
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 272 AGCGGGAGTGGGAGAACTACATCACTCTCTCT - GGAGAGGCGGAGTGGAGGGCTGCTGCC 330
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 248 AGCGGGAGTGGGAGAACTACATCACTCTCTCTGGAGAGGGGAGTGGAGGGCTGCTGCC 307
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 331 TG 332


```

Db 308 TG 309
||
RESULT 13
AA747293 157 bp mRNA EST 16-JAN-1998
LOCUS aa747293
DEFINITION nx88a10.s1 NCI-CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1269306
similar to TR:Q64906 Q64906 SIMILAR TO GENBANK ACCESSION NUMBER
L2608..; mRNA sequence.
ACCESSION AA747293
NID 92787251
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 157)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Martin, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
Location/Qualifiers
1..157
/organism="Homo sapiens"
/Note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Martin (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1269306"
/clone_lib="NCI-CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/lab_host="DH10B"

BASE COUNT 33 a 43 c 54 g 26 t 1 others
ORIGIN

Query Match 6.7%; Score 127.2; DB 26; Length 157;
Best Local Similarity 93.5%; Pred. No. 7.8e-24;
Matches 143; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 211 GCCAGACGATCTGTCGGCAGGTGAATATCGGCTCCACAAGGGTCTGCTGTGAT 270
|||||
Db 5 GGCAAGACGATCTGTCGGCAGGTGAATATCGGCTCCACAAGGGTCTGCTGTGAT 64
|||||

Qy 271 AAGCGGGACTGCGAGAACTACATCATCTCTCTGGAGAGCGGAGTGAGGGGCTGCTGCC 330
|||||

```

```

Db 65 AAGCGGGACTGCGAGAACTACATCATCTCTCTGGAGAGCGGAGTGAGGGGNCCTGCTGCC 124
Qy 331 TGTGGCACCACGCGCGCGCACCACCCAGCTGTGG 363
|||||
Db 125 TGTGG-ACCAACGCGCGCGCACCACCCAGCTGTG 156

RESULT 14
AA023538 492 bp mRNA EST 21-JAN-1997
LOCUS aa023538/c
DEFINITION mh75d07.r1 Soares mouse placenta 4NBMP13.5 14.5 Mus musculus cDNA
clone 456781 5' similar to PIR:JH0798 JH0798 fasciclin IV precursor
- American bird grasshopper ;, mRNA sequence.
ACCESSION AA023538
NID 91487435
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 492)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicinep
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:273669
Possible reversed clone: similarity on wrong strand
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 473.
Location/Qualifiers
1..492
/organism="Mus musculus"
/strain="C57BL/6J"
/Note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGGAAATTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:i0090"
/clone_lib="Soares mouse placenta 4NBMP13.5 14.5"
/clone="456781"
/sex="unknown"
/dev_stage="adult"
/lab_host="DH10B"
/lab_host="DH10B"

BASE COUNT 135 a 121 c 111 g 124 t 1 others
ORIGIN

Query Match 3.0%; Score 57.4; DB 15; Length 492;
Best Local Similarity 52.8%; Pred. No. 2.9e-05;
Matches 124; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 553 ACTCTGTCGAGACCCACAGTTCATCAAGCCACCATCGTGCACCAGACCGCTTAC 612
|||||

```

WORLD

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:02:47 1999; MasPar time 7.55 Seconds
Tabular output not generated. 72.281 Million cell updates/sec

Title: >US-09-041-236-2
Description: (1-10) from US09041236.pep (1 of 45)
Perfect Score: 74
Sequence: 1 LLLLLWAAA 10

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp Vertebrate 14:sp_virus

Statistics: Mean 25.106; Variance 52.424; scale 0.479

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	74	100.0	666	4	SEMOPHORIN L.	2.96e+01
2	70	94.6	592	13	7-TRANSMEMBRANE PROTEIN	1.12e+00
3	67	90.5	190	7	MHC-CLASS I (FRAGMENT)	2.95e+00
4	67	90.5	302	7	LEUKOCYTE ANTIGEN B (H	2.95e+00
5	67	90.5	322	7	HLA-B*37 (FRAGMENT)	2.95e+00
6	67	90.5	341	7	MHC CLASS I HISTOCOMPA	2.95e+00
7	67	90.5	354	7	CLASS I HISTOCOMPATIBI	2.95e+00
8	67	90.5	355	7	HLA-B ALPHA-CHAIN (FRA	2.95e+00
9	67	90.5	359	7	HLA-B*1802 (FRAGMENT)	2.95e+00
10	67	90.5	362	7	HLA-B*1802	2.95e+00
11	67	90.5	362	7	MHC CLASS I HLA-B13 SU	2.95e+00
12	67	90.5	362	7	HLA-B ALPHA CHAIN	2.95e+00
13	67	90.5	362	7	MHC ANTIGEN HLA-B*1303	2.95e+00
14	67	90.5	362	7	MHC CLASS I ANTIGEN	2.95e+00
15	67	90.5	362	7	LYMPHOCYTE ANTIGEN PRE	2.95e+00
16	67	90.5	362	7	LYMPHOCYTE ANTIGEN	2.95e+00
17	67	90.5	362	7	HLA-B PROTEIN	2.95e+00
18	67	90.5	362	7	MHC CLASS I HLA-B44BO	2.95e+00
19	67	90.5	362	7	MHC CLASS I ANTIGEN	2.95e+00
20	67	90.5	362	7	MHC CLASS I ALPHA CHAI	2.95e+00

21	67	90.5	362	7	Q29850	HLA-B ALPHA CHAIN B*44	2.95e+00
22	67	90.5	362	7	Q29705	MHC CLASS I ANTIGEN HL	2.95e+00
23	67	90.5	362	7	Q78180	HLA CLASS I ANTIGEN HL	2.95e+00
24	67	90.5	362	7	Q29849	LYMPHOCYTE ANTIGEN	2.95e+00
25	67	90.5	362	7	Q29846	HUMAN LYMPHOCYTE ANTIG	2.95e+00
26	65	87.8	1838	4	Q15094	PRO-ALPHA-1 TYPE V COL	5.59e+00
27	64	86.5	362	7	Q79523	MHC CLASS I HISTOCOMPA	7.66e+00
28	64	86.5	362	7	Q29718	MHC CLASS I ANTIGEN	7.66e+00
29	64	86.5	362	7	Q79489	HLA-B*55.2	7.66e+00
30	64	86.5	362	7	Q79490	HLA-B*22N	7.66e+00
31	64	86.5	362	7	Q29940	MHC HLA-B TRANSMEMBRAN	7.66e+00
32	64	86.5	1840	11	Q60467	PRO-ALPHA-1 TYPE V COL	7.66e+00
33	61	82.4	35	4	Q13380	RNA BINDING MOTIF (RBM	1.95e+01
34	61	82.4	45	4	Q13379	RNA BINDING MOTIF (RBM	1.95e+01
35	60	81.1	60	4	Q14478	BETA-GLOBIN	2.64e+01
36	60	81.1	83	14	Q84277	ES PROTEIN	2.64e+01
37	60	81.1	177	11	Q88186	PLATELET GLYCOPROTEIN	2.64e+01
38	60	81.1	202	7	Q78050	MHC CLASS I ANTIGEN HL	2.64e+01
39	60	81.1	202	7	Q78052	MHC CLASS I ANTIGEN HL	2.64e+01
40	60	81.1	322	7	Q19626	HLA-B*35 (FRAGMENT)	2.64e+01
41	60	81.1	339	7	Q95353	MHC CLASS I HLA-B*62 (2.64e+01
42	60	81.1	362	7	Q29851	GERMLINE (HLA-B) (FRAG	2.64e+01
43	60	81.1	362	7	Q78218	LYMPHOCYTE ANTIGEN	2.64e+01
44	60	81.1	362	7	Q19755	MHC CLASS I ANTIGEN HL	2.64e+01
45	60	81.1	445	2	Q55937	NICKEL RESISTANCE	2.64e+01

ALIGNMENTS

RESULT 1							
ID	Q75326	PRELIMINARY;	PRT;	666	AA.		
AC	Q75326;						
DT	01-NOV-1998	(TREMREL. 08, CREATED)					
DT	01-NOV-1998	(TREMREL. 08, LAST SEQUENCE UPDATE)					
DT	01-NOV-1998	(TREMREL. 08, LAST ANNOTATION UPDATE)					
DE	SEMOPHORIN L.						
GN	SEMAL						
OS	HOMO SAPIENS (HUMAN)						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;						
OC	CATARRHINI; HOMINIDAE; HOMO.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 98389619.						
RA	LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;						
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA						
RT	viruses."						
RL	GENOMICS 51:340-350(1998).						
DR	EMBL; AF030698; G3523115; -						
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;						

Query Match 100.0%; Score 74; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 2.96e+01;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	33	LLLLWAAA	42				
QY	1	LLLLWAAA	10				
RESULT 2							
ID	Q57328	PRELIMINARY;	PRT;	592	AA.		
AC	Q57328;						
DT	01-JUN-1998	(TREMREL. 06, CREATED)					
DT	01-JUN-1998	(TREMREL. 06, LAST SEQUENCE UPDATE)					
DT	01-JUN-1998	(TREMREL. 06, LAST ANNOTATION UPDATE)					
DE	7-TRANSMEMBRANE PROTEIN FRIZZLED-1.						
GN	FZ-1						
OS	GALLUS GALLUS (CHICKEN)						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;						
OC	NEOGATHAE; GALLIFORMES; PHASIANIDAE; GALLUS.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	TISSUE-LIMB BUD;						

RA KENGAKU M., TWOMBLY V., TABIN C.;
 RL COLD SPRING HARB. SYMP. QUANT. BIOL. 0:0-0(1997).
 DR EMBL; AF031830; G2655274; -.
 KW TRANSMEMBRANE.
 SQ SEQUENCE 592 AA; 65490 MW; CD66EC83 CRC32;

Query Match 94.6%; Score 70; DB 13; Length 592;
 Best Local Similarity 100.0%; Pred. No. 1.12e+00;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 35 LLLLLWAAA 43
 |||||

Qy 1 LLLLLWAAA 9

RESULT 3
 ID Q29856 PRELIMINARY; PRT; 190 AA.

AC Q29856;
 DT 01-NOV-1996 (TREMREL. 01, CREATED)
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE MHC-CLASS I (FRAGMENT).
 GN HLA-B.

OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 [1]

RN SEQUENCE FROM N.A.
 RX MEDLINE; 95278975.

RA YAO Z., LATTERMANN A., VOLGER A., MCNICHOLAS A., MUELLER-ECKHARDT G.,
 ALBERT E.;

RL "A new HLA-B*44 subtype, B*4406, differing in exon 2."
 RT IMMUNOGENETICS 41:387-387(1995).

DR EMBL; X83401; E198011; -.

DR EMBL; X83402; E198011; JOINED.

DR EMBL; X83403; E198011; JOINED.

DR PFAM; PF00129; MHC_I; 1.

KW MHC.

FT NON_TER 190 190

SQ SEQUENCE 190 AA; 21223 MW; BFC0430E CRC32;

Query Match 90.5%; Score 67; DB 7; Length 190;
 Best Local Similarity 80.0%; Pred. No. 2.95e+00;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18

Qy 1 LLLLLWAAA 10

RESULT 4
 ID Q95392 PRELIMINARY; PRT; 302 AA.

AC Q95392;
 DT 01-FEB-1997 (TREMREL. 02, CREATED)
 DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE LEUKOCYTE ANTIGEN B (HLA-B*47 VARIANT) (FRAGMENT).
 GN HLA-B.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 [1]

RN SEQUENCE FROM N.A.

RC TISSUE-PERIPHERAL BLOOD;

RA FISCHER G.F., FAE I., LEITNER D., MAYR W.R.;

RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; J09118; E280547; -.

DR PROSITE; PS00290; IG_MHC; 1.

DR PFAM; PF00047; Ig; 1.

DR PFAM; PF00129; MHC_I; 1.

KW MHC.

FT NON_TER 302 302

SQ SEQUENCE 302 AA; 34828 MW; D395D4B7 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 302;
 Best Local Similarity 80.0%; Pred. No. 2.95e+00;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18

Qy 1 LLLLLWAAA 10

RESULT 5
 ID O19627 PRELIMINARY; PRT; 322 AA.

AC O19627;
 DT 01-JAN-1998 (TREMREL. 05, CREATED)
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)

DE HLA-B*37 (FRAGMENT).

GN B-3701.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;

OC CATARRHINI; HOMINIDAE; HOMO.
 [1]

RN SEQUENCE FROM N.A.

RA HURLEY C.K., BEI M., RODRIGUEZ S., JOHNSON A.;

RL SUBMITTED (JUN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; U11267; G511786; -.

DR PROSITE; PS00290; IG_MHC; 1.

DR PFAM; PF00047; Ig; 1.

DR PFAM; PF00129; MHC_I; 1.

KW MHC.

FT NON_TER 322 322

SQ SEQUENCE 322 AA; 36626 MW; DF3B7744 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 322;
 Best Local Similarity 80.0%; Pred. No. 2.95e+00;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18

Qy 1 LLLLLWAAA 10

RESULT 6
 ID O19589 PRELIMINARY; PRT; 341 AA.

AC O19589;
 DT 01-JAN-1998 (TREMREL. 05, CREATED)
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE MHC CLASS I HISTOCOMPATIBILITY ANTIGEN HLA-B (FRAGMENT).
 GN HLA-B.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;

OC CATARRHINI; HOMINIDAE; HOMO.
 [1]

RN SEQUENCE FROM N.A.

RA SANTOS S., BALAS A., VICARIO J.L.;

RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF022783; G2564648; -.

DR PROSITE; PS00290; IG_MHC; 1.

DR PFAM; PF00047; Ig; 1.

DR PFAM; PF00129; MHC_I; 1.

KW MHC.

FT NON_TER 341 341

SQ SEQUENCE 341 AA; 38416 MW; E97896CA CRC32;

Query Match 90.5%; Score 67; DB 7; Length 341;
 Best Local Similarity 80.0%; Pred. No. 2.95e+00;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18

Qy 1 LLLLLWAAA 10

```

RESULT 7
ID Q95530 PRELIMINARY; PRT: 354 AA.
AC Q95530:
DT 01-FEB-1997 (TREMREL. 02, CREATED)
DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE CLASS I HISTOCOMPATIBILITY ANTIGEN (FRAGMENT).
GN HLA-B.
OS PAN TROGLODYTES (CHIMPANZEE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; PAN.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TEPPE.
RX MEDLINE: 94286544.
RA MCADAM S.N., BOYSON J.E., LIU X., GARBER T.L., HUGHES A.L.,
RA BONTROP R.E., WATKINS D.I.;
RT "A uniquely high level of recombination at the HLA-B locus.";
RL PROC. NATL. ACAD. SCI. U.S.A. 91:5893-5897(1994).
DR EMBL: U05581; G454780;
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; Ig; 1.
DR PFAM: PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1
SQ SEQUENCE 354 AA; 39555 MW; 0688F963 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 354;
Best Local Similarity 80.0%; Pred.No. 2.95e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1 LLLLLWGAVA 10
| | | | | | | |
QY 1 LLLLLWAAAA 10

RESULT 8
ID Q29853 PRELIMINARY; PRT: 355 AA.
AC Q29853:
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE HLA-B ALPHA-CHAIN (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95176328.
RA GAUCHAT-FEISS D., BREUR-VRIESENDORP B.S., RUFER N., JEANNET M.,
RA ROOSNEK E., TIERCY J.M.;
RT "Sequencing of a novel functional HLA-B44 subtype differing in two
RT residues in the alpha 2 domain.";
RL TISSUE ANTIGENS 44:261-264(1994).
DR EMBL: X75953; G791008;
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; Ig; 1.
DR PFAM: PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 355
SQ SEQUENCE 355 AA; 39845 MW; EF81934E CRC32;

Query Match 90.5%; Score 67; DB 7; Length 355;
Best Local Similarity 80.0%; Pred.No. 2.95e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
| | | | | | | |
QY 1 LLLLLWAAAA 10

RESULT 9
ID Q29934 PRELIMINARY; PRT: 359 AA.
AC Q29934:
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE HLA-B27 (FRAGMENT).
GN HLA-B.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86149317.
RA SZOTS H., RIETHMULLER G., WEISS E., MEO T.;
RT "Complete sequence of HLA-B27 cDNA identified through the
RT characterization of structural markers unique to the HLA-A, -B, and
RT -C allelic series.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:1428-1432(1986).
DR EMBL: M12678; G386882;
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; Ig; 1.
DR PFAM: PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1
SQ SEQUENCE 359 AA; 40042 MW; 069F7E64 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 359;
Best Local Similarity 80.0%; Pred.No. 2.95e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 6 LLLLLWGAVA 15
| | | | | | | |
QY 1 LLLLLWAAAA 10

RESULT 10
ID Q29845 PRELIMINARY; PRT: 362 AA.
AC Q29845:
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE HLA-B*1802.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP PARTIAL SEQUENCE FROM N.A.
RX TISSUE-BLOOD;
RA LIN L.;
RL SUBMITTED (NOV-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 115-131 FROM N.A.
RC TISSUE-BLOOD;
RX MEDLINE: 95270437.
RA LIN L., TOKUNAGA K., ISHIKAWA Y., KASHIWASE K., NAKAJIMA F.,
RA NISHIMURA M., KUWATA S., AKAZA T., TADOKORO K., SHIBATA Y., JUJI T.;
RT "A new B18 sequence (B*1802) from Asian individuals.";
RL HUM. IMMUNOL. 42:23-26(1995).
DR EMBL: D25275; D1005505;
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; Ig; 1.
DR PFAM: PF00129; MHC_I; 1.
SQ SEQUENCE 362 AA; 40232 MW; E2645034 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 362;
Best Local Similarity 80.0%; Pred.No. 2.95e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
| | | | | | | |
QY 1 LLLLLWAAAA 10

```

```
RESULT 11
ID Q29933 PRELIMINARY; PRT; 362 AA.
AC Q29933; O02956;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MHC CLASS I HLA-B*13 SUBTYPE B13.1 (ORIENTAL) (FRAGMENT).
GN HLA-B.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89122134.
RA KATO K., DUPONT B., YANG S.Y.;
RT "Localization of nucleotide sequence which determines Mongoloid
RT subtype of HLA-B13";
RL IMMUNOGENETICS 29:117-120(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RX MEDLINE; 96053518.
RA LIN L., TOKUNAGA K., NAKAJIMA F., ISHIKAWA Y., KASHIWASE K.,
RA TANAKA H., KUWATA S., SIDELTSEVA E., AKAZA T., TADOKORO K.,
RA SHIBATA Y., CHANDANAYINGYONG D., JUJI T.;
RT "Both HLA-B*1301 and B*1302 exist in Asian populations and are
RT associated with different haplotypes.";
RL HUM. IMMUNOL. 43:51-56(1995).
DR EMBL; M24075; G386884; -
DR EMBL; D50290; G1682952; -
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 362 362
SQ SEQUENCE 362 AA; 40468 MW; 971DIFEE CRC32;

Query Match 90.5%; Score 67; DB 7; Length 362;
Best Local Similarity 80.0%; Pred. No. 2.95e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
| | | | | : | |
Qy 1 LLLLLWAAAA 10

RESULT 12
ID Q29842 PRELIMINARY; PRT; 362 AA.
AC Q29842;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HLA-B ALPHA CHAIN.
GN HLA-B.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX HERRERO M.J.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X84725; G683575; -
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT CHAIN 25 362 HLA-B ALPHA CHAIN.
SQ SEQUENCE 362 AA; 40592 MW; 67CE0F02 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 362;
Best Local Similarity 80.0%; Pred. No. 2.95e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Db 9 LLLLLWGAVA 18
| | | | | : | |
Qy 1 LLLLLWAAAA 10

RESULT 13
ID Q29661 PRELIMINARY; PRT; 362 AA.
AC Q29661;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MHC ANTIGEN HLA-B*1303 PRECURSOR.
GN HLA-B*1303.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RX MEDLINE; 96232969.
RA BALAS A., GARCIA-SANCHEZ F., VICARIO J.L.;
RT "HLA-B*1303; a new example of poor correlation between serology and
RT structure.";
RL HUM. IMMUNOL. 45:32-36(1996).
DR EMBL; U14943; G608462; -
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW SIGNAL; MHC.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 362 MHC ANTIGEN HLA-B*1303.
SQ SEQUENCE 362 AA; 40501 MW; 9277DEFB CRC32;

Query Match 90.5%; Score 67; DB 7; Length 362;
Best Local Similarity 80.0%; Pred. No. 2.95e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
| | | | | : | |
Qy 1 LLLLLWAAAA 10

RESULT 14
ID Q29693 PRELIMINARY; PRT; 362 AA.
AC Q29693;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MHC CLASS I ANTIGEN.
GN HLA-B.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96134006.
RA RUDWALEIT M., BOWNESS P., WORDSWORTH P.;
RT "The nucleotide sequence of HLA-B*2704 reveals a new amino acid
RT substitution in exon 4 which is also present in HLA-B*2706.";
RL IMMUNOGENETICS 43:160-162(1996).
DR EMBL; U27608; G1276485; -
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
SQ SEQUENCE 362 AA; 40416 MW; 850432B4 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 362;
Best Local Similarity 80.0%; Pred. No. 2.95e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
| | | | | : | |
```

QY 1 LLLLLWAAA 10

RESULT 15
 ID Q30173 PRELIMINARY; PRT; 362 AA.
 AC Q30173;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE LYMPHOCYTE ANTIGEN PRECURSOR.
 GN HLA-B.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93056529.
 RA HILDEBRAND W.H., MADRIGAL J.A., BELICH M.P., ZENMOUR J., WARD F.E.,
 RA WILLIAMS R.C., PARHAM P.;
 RT "Serologic cross-reactivities poorly reflect allelic relationships in
 RT the HLA-B12 and HLA-B21 groups. Dominant epitopes of the alpha 2
 RT helix.";
 RL J. IMMUNOL. 149:3563-3568(1992).
 DR EMBL; M84694; G188475; -.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PFAM; PF00047; ig; 1.
 DR PFAM; PF00129; MHC_I; 1.
 KW SIGNAL; MHC.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 362 LYMPHOCYTE ANTIGEN.
 SQ SEQUENCE 362 AA; 40519 MW; 91EDFB3 CRC32;
 Query Match 90.5%; Score 67; DB 7; Length 362;
 Best Local Similarity 80.0%; Pred. No. 2.95e+00;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 9 LLLLLWGAVA 18
 QY 1 LLLLLWAAA 10

Search completed: Thu Jul 8 18:02:59 1999
 Job time : 12 secs.

THIS PAGE BLANK (USPTO)

工
L
Z
U
回
Z

DT 12-JAN-1995 (first entry)
DE Human collagen (Type V).
KW Human collagen; alpha 1; V type collagen; placental mRNA.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..38
FT region 444..538
FT binding_site 645..647
FT /note= "contains (Gly-X-Y) repeats"
FT /label= RGD
FT /note= "cell adhesion motif"
FT binding_site 663..665
FT /label= RGD
FT /note= "cell adhesion motif"
FT domain 837..933
FT /label= heparin_binding_domain
FT region 1573..1838
FT /label= C-terminal_region
FT /note= "contains 8 Cys residues"
PN J06105687-A.
PD 19-APR-1994.
PF 27-DEC-1991; 358300.
PR 27-DEC-1991; JP-358300.
PA (TAKI) TAKARA SHUZO CO LTD.
DR WPI: 94-163129/20.
DR N-PSDB: Q64556.
PT Human collagen V-type gene - is used for diagnosis of human collagen V-type related diseases
PT Claim 1; Page 6-14; 19pp; Japanese.
CC This amino acid sequence of type V collagen contains several distinct domains including a region comprising repeated (Gly-X-Y) motifs and a central domain containing two copies of the Arg-Gly-Asp cell adhesion motif. The cDNA sequence encoding type V collagen was isolated from a human placental library and will be useful for diagnosis of diseases related to type V collagen.
CC Sequence 1838 AA;
SQ
Query Match 87.8%; Score 65; DB 10; Length 1838;
Best Local Similarity 70.0%; Pred. No. 1.26e+02;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
DB 24 lllllwapp 33
QY 1 LLLLLWAAA 10
RESULT 3
ID W23244 standard; Protein: 406 AA.
AC W23244;
DT 19-MAR-1998 (first entry)
DE Human renin.
KW Transgenic rat; human; renin; angiotensinogen; model;
KW hypertension; screening; treatment; conversion enzyme; ACE;
KW inhibitor.
OS Homo sapiens.
PN WO9728684-A1.
PD 14-AUG-1997; J00289.
PF 06-FEB-1997; J00289.
PR 07-FEB-1996; JP-021034.
PA (FUJI) FUJISAWA PHARM CO LTD.
PI Ishikawa H, Kobayashi M, Mukumoto S, Satoh S;
DR WPI: 97-448318/41.
DR N-PSDB: T84705.
PT Transgenic hypertensive rat model for study of hypertension - contains DNA encoding human renin and angiotensinogen linked to promoters for high expression of these genes in rats
PS Example: Pages 26-29; 44pp; Japanese.
CC The present sequence can be used in the preparation of a novel transgenic hypertensive rat, comprising DNA encoding human renin and angiotensinogen, linked to promoters for high expression of these genes in rats. The rat can be used as a model for the study of hypertension, and to screen possible candidate drugs for the

CC treatment of hypertensive disorders, e.g. renin and angiotensinogen conversion enzyme (ACE) inhibitors.
SQ Sequence 406 AA;
Query Match 82.4%; Score 61; DB 26; Length 406;
Best Local Similarity 75.0%; Pred. No. 2.30e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
DB 12 llllllwws 19
QY 1 LLLLLWAA 8
RESULT 4
ID P50135 standard; Protein: 406 AA.
AC P50135;
DT 20-NOV-1991 (first entry)
DE Sequence of pre-pro-renin.
KW Blood pressure control; renin; pro-renin; pre-pro-renin.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..20
FT /label= signal
FT peptide 21..66
FT /label= pro-sequence
FT protein 67..406
PN EP-135347-A.
PD 27-MAR-1985.
PF 08-AUG-1984; 305412.
PR 12-AUG-1983; JP-147472.
PA (MURA/) MURAKAMI K.
PA (DAUC) DAIICHI SEIYAKU KK.
PI Murakami K, Nakanishi S;
DR WPI: 85-076060/13.
DR N-PSDB: N50180.
PT Human renin poly:peptide expressed by Escherichia coli - with recombinant DNA sequence for transformation
PS Disclosure; Fig 2; 24pp; English.
CC The inventors claim the pre-pro-sequence, the pro-sequence and the mature sequence of renin, and the cDNA encoding these SQs, as depicted in Fig 2. The cDNA may be obt'd. by sepn. of intracytoplasmic DNA from a kidney fragment isolated from a patient showing high serum renin. An oligo(dT) cellulose column is used to obtain mRNA.
CC Sequence 406 AA;
SQ

Query Match 82.4%; Score 61; DB 3; Length 406;
Best Local Similarity 75.0%; Pred. No. 2.30e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
DB 12 llllllwws 19
QY 1 LLLLLWAA 8
RESULT 5
ID R28371 standard; Protein: 83 AA.
AC R28371;
DT 20-APR-1993 (first entry)
DE Papillomavirus E5 protein.
KW Human; papillomavirus; E5 gene; keratinocyte; epithelial; cell line;
KW bovine; papillomavirus; BPV; fibroblast; fibropapillomavirus; pig;
KW HPV; polymerase chain reaction; PCR; HPV16DNA; mouse; rat; rabbit;
KW goat; neoplastic; proliferation; HPV16; warts; papillomas; cervical; dysplasia; carcinoma.
OS Homo sapiens.
PN WO9220784-A.
PD 26-NOV-1992.
PF 23-AUG-1991; U06039.
PR 16-MAY-1991; US-701413.
PA (UYIA) UNIV YALE.
PI Dimaio DC, Dotto G;
DR WPI: 92-415763/50.

DR N-PSDB; Q31691.
PT Cell line expressing human papilloma virus E5 gene - for
PT screening of drugs that inhibit action of E5 gene, and to
PT determine role of E5 gene expression in disease
PS Claim 8; Page 34; 42pp; English.
CC The sequence given is encoded by the human papillomavirus E5 gene
CC which was isolated from a keratinocyte or epithelial cell line. The
CC E5 gene has been cited as being involved in the ability of bovine
CC papillomavirus (BPV) to transform fibroblasts in vitro. It is well
CC conserved among various fibropapillomaviruses. The human
CC papillomavirus (HPV) E5 gene was amplified using polymerase chain
CC reaction from the plasmid HPV16DNA. The amplified sequence can be
CC introduced into a keratinocyte or epithelial cell line from an animal
CC species such as mouse, rat, rabbit, pig or goat. The HPV E5 gene
CC is a possible contributor to the neoplastic proliferation induced by
CC HPV16 in vivo. Therefore it is considered that HPV E5 expression
CC will play a role in human diseases caused by these viruses including
CC warts (papillomas), cervical dysplasia and probably carcinoma.
CC Sequence 83 AA;
SQ
Query Match 81.1%; Score 60; DB 6; Length 83;
Best Local Similarity 70.0%; Pred. No. 2.67e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 45 lvvllwitaa 54
:|||||:
QY 1 LLLLLWAAAA 10
RESULT 6
ID R02234 standard; protein; 83 AA.
AC R02234;
DE 28-AUG-1990 (first entry)
DE Sequence of the E5 oncogene region of human papillomavirus type 16
DE (HPV16)
KW Human papillomavirus type 16; E5 oncogene; transactivating protein;
KW viral replication antagonist; therapeutic agent.
OS Human papillomavirus type 16.
PN W08912461-A.
PD 28-DEC-1989.
PF 01-JUN-1989; U02404.
PR 22-MAY-1989; US-352688, US-207393.
PA (USYL-) St Louis Univ.
PA Green M, Loewenstein PM;
DR WPI: 90-022386/03.
PT New antigens of viral, esp. HIV, replication -
PT are fragments of trans-activating protein, opt. contg. aminoacid
PT substitution
PS Disclosure: 68pp; English.
CC The invention provides for a method of preparing peptide antagonists of
CC viral transactivating proteins (TAPs) by producing peptide fragments of
CC the TAP and then selecting those which show antagonistic activity.
CC Alternatively, peptides contg. an active domain of TAP (or the full
CC length protein) are prep'd., and one or more AA substn(s). is made to
CC eliminate transactivating activity without destroying binding ability.
CC The antagonists are useful as therapeutic agents in the control of viral
CC infection and viral pathogenicity. The cellular DNA synthesis-inducing
CC protein encoded by the bovine papillomavirus type 1 (BPV1) E5 oncogene
CC is one of the viral systems which the description of the application
CC of the invention focuses on. E5 is conserved among animal papillomavirus
CC type 1. Moreover, E5 is also conserved among two strong candidates for
CC causative agents of cervical carcinoma and other urogenital malignancies
CC (human papillomavirus type 16 (HPV16) and HPV18), as well as one of the
CC causative agents of venereal condylomas and juvenile laryngeal
CC papillomatosis.
CC Sequence 83 AA;
SQ
Query Match 81.1%; Score 60; DB 1; Length 83;
Best Local Similarity 70.0%; Pred. No. 2.67e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 45 lvvllwitaa 54
:|||||:
QY 1 LLLLLWAAAA 10

QY 1 LLLLLWAAAA 10
RESULT 7
ID R03142 standard; protein; 362 AA.
AC R03142;
DE 19-MAR-1991 (first entry)
DE Sequence of HLA-Bw52 antigen.
KW Probe: HLA class I DNA; immunogen.
OS Homo sapiens.
PN EP-354580-A.
PD 14-FEB-1990.
PF 10-AUG-1989.
PR 11-AUG-1988; JP-200758.
PA (OLYU) Olympus Optical Co., Ltd.
PI Kano K, Takiguchi;
DR WPI: 90-046289/07.
PT New DNA for class I human leucocyte antigens and derived probes and
PT transformed cells, useful for DNA typing, as immunogens etc.
PS Disclosure: Page 13; 23pp; English.
CC The HLA class I DNA can be used as a source of probes for use in DNA
CC typing. Transformed cells, which are useful as immunogens, can be
CC obtained by introducing these DNAs into eucaryotic cells.
CC Sequence 362 AA;
SQ
Query Match 81.1%; Score 60; DB 2; Length 362;
Best Local Similarity 70.0%; Pred. No. 2.67e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 9 vllllwgava 18
:|||||:
QY 1 LLLLLWAAAA 10
RESULT 8
ID R12464 standard; Protein; 362 AA.
AC R12464;
DE 29-AUG-1991 (first entry)
DE HLA-B35 antigen.
KW Human leukocyte antigen; probe; major histocompatibility complex;
KW MHC; class I.
OS Homo sapiens.
PN J03112486-A.
PD 14-MAY-1991.
PF 22-SEP-1989; 247697.
PR 22-SEP-1989; JP-247697.
PA (OLYU) OLYMPUS OPTICAL KK.
DR WPI: 91-182991/25.
DR N-PSDB; Q12115.
PT HLA-B35 gene - used in DNA probe and transformant cells for
PT immunising animals, for developing monoclonal antibody.
PS Claim 1; Page 1; 11pp; Japanese.
CC Probes comprising part of the sequence encoding this sequence can
CC be used to identify class I genes. The DNA can be expressed for
CC immunisation of animals and prodn. of monoclonal antibodies
CC specific for the HLA-B35 antigen. See also J03112485 and J03112487.
CC Sequence 362 AA;
SQ
Query Match 81.1%; Score 60; DB 3; Length 362;
Best Local Similarity 70.0%; Pred. No. 2.67e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 9 vllllwgava 18
:|||||:
QY 1 LLLLLWAAAA 10
RESULT 9
ID R12463 standard; Protein; 362 AA.
AC R12463;
DE 29-AUG-1991 (first entry)
DE HLA-Bw53 exon.
KW Human leukocyte antigen; probe; major histocompatibility complex;
KW

```

KW MHC; class I.
OS Homo sapiens.
PD J03112487-A.
PN 14-MAY-1991.
PF 22-SEP-1989; 247697.
PR 22-SEP-1989; JP-247697.
PA (OLYU ) OLYMPUS OPTICAL KK.
DR WPI: 91-182991/25.
DR N-PSDB; Q12114.
PT HLA-Bw53 gene, DNA probe and transformant cells - used for
PT immunisation, identifying specificity of antiserum etc.
PS Claim 2: Page 1; 11pp; Japanese.
CC Probes comprising part of the sequence encoding the protein can be
CC used to identify Class I genes. The DNA can be expressed for
CC immunisation of animals and prodn. of monoclonal antibodies
CC specific for the HLA-Bw53 antigen. See also J03112485 and
CC J03112486.
SQ Sequence 362 AA;

Query Match 81.1%; Score 60; DB 3; Length 362;
Best Local Similarity 70.0%; Pred. No. 2.67e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 vllllwgava 18
QY 1 LLLLLWAAAA 10

RESULT 10
ID R03144 standard; protein; 362 AA.
AC R03144;
DE 19-MAR-1991 (first entry)
DE Sequence of HLA-B51 antigen.
KW Probe: HLA class I DNA; immunogen.
OS Homo sapiens.
PN EP-354580-A.
PD 14-FEB-1990.
PF 10-AUG-1989.
PR 11-AUG-1988; JP-200758.
PA (OLYU) Olympus Optical Co., Ltd.
PI Kano K, Takiguchi;
DR WPI: 90-046289/07.
PT New DNA for class I human leucocyte antigens and derived probes and
PT transformed cells, useful for DNA typing, as immunogens etc.
PS Disclosure: Pages 12-13; 23pp; English.
CC The HLA class I DNA can be used as a source of probes for use in DNA
CC typing. Transformed cells, which are useful as immunogens, can be
CC obtained by introducing these DNAs into eucaryotic cells.
SQ Sequence 362 AA;

Query Match 81.1%; Score 60; DB 2; Length 362;
Best Local Similarity 70.0%; Pred. No. 2.67e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 vllllwgava 18
QY 1 LLLLLWAAAA 10

RESULT 11
ID W42034 standard; Protein; 109 AA.
AC W42034;
DE Mouse preprocortistatin protein.
DE Mouse preprocortistatin protein.
KW Mouse preprocortistatin; N-terminal signal peptide; procortistatin;
KW cortistatin-29; cortistatin-14; agonist; antibody; inhibition; sleep;
KW somatostatin; neuronal electrical activity; cerebral cortex; antagonist;
KW acetylcholine; primer; PCR; immunocassay; hybridisation; amplification.
OS Mus sp.
FH Key Location/Qualifiers
FT Peptide 1..27
FT Protein /note= "signal peptide"
28..109

```

```

FT /note= "mature peptide"
FT Cleavage_site 79..80
FT /note= "cleaves to give a putative 29-residue
FT precursor"
FT Cleavage_site 94..95
FT /note= "cleaves to give rise to a putative 13 and
FT 14 residue peptide"
FT WO3743417-A1.
PD 20-NOV-1997.
PF 15-MAY-1997; U08481.
PR 15-MAY-1996; US-648322.
PA (SCRI ) SCRIPPS RES INST.
PI De Lecea L, Henriksen SJ, Siggins GR, Sutcliffe JG;
DR WPI: 98-008886/01.
DR N-PSDB; V09156.
PT New cortistatin peptide(s) - used to modulate sleep, detect
PT mutation(s) and screen for drugs
PT Claim 1; Page 104; 128pp; English.
CC This amino acid sequence is the novel mouse preprocortistatin protein.
CC cleavage at the N-terminal signal peptide site leads to the formation
CC of procortistatin, which is processed at any of the tandem basic amino
CC acid pairs Lys-Ser, or Lys-Lys to produce the mature cortistatin
CC proteins referred as mouse cortistatin-29, and cortistatin-14. The
CC purified cortistatin, and its agonists, are used to induce sleep while
CC its receptor antagonists (particularly antibodies) is used to inhibit
CC sleep. Although cortistatin is structurally similar to somatostatin,
CC it is able to depress neuronal electrical activity, induce low frequency
CC waves in the cerebral cortex, antagonise acetylcholine and therefore
CC enhance slow-wave sleep. The antibodies, and oligonucleotide primers,
CC are used in usual immunoassays and hybridisation/amplification assays to
CC detect or quantify cortistatin (including that administered
CC therapeutically) or its nucleic acid. Oligonucleotides, e.g. antisense
CC molecules, are used in vivo to alter cortistatin gene expression.
CC Detection of a mutation in the cortistatin gene may provide diagnosis of
CC sleep-related or neuronal depression-related disorders or diseases of
CC the brain.
SQ Sequence 109 AA;

Query Match 79.7%; Score 59; DB 29; Length 109;
Best Local Similarity 77.8%; Pred. No. 3.09e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 17 lllllwgava 25
QY 2 LLLLLWAAAA 10

RESULT 12
ID R04386 standard; protein; 116 AA.
AC R04386;
DE 13-SEP-1990 (first entry)
DE Colon Cancer monoclonal antibody CC49 light chain variable region
DE chimeric antibodies; TAG72; light chain variable region;
KW heavy chain variable region; ss; colon cancer.
OS synthetic.
FH Key Location/Qualifiers
FT Peptide 1..20
FT /label= leader peptide
FT region 21..104
FT /label= variable region
FT region 105..116
FT /label= J5 segment
FT EP-365997-A.
PD 02-MAY-1990.
PF 18-OCT-1989; 119361.
PR 19-OCT-1988; US-259943.
PA (DOMC) Dow Chemical Co.
PI Mezes P, Gourlie B, Rixon M;
DR WPI: 90-133521/18.
DR N-NSDB; Q04253.
PT Chimeric antibodies against TAG72 and conjugate to provide imaging
PT markers and therapeutic tools
PS Disclosure; pp; English.

```

CC The polypeptide forms part of a chimera. The other components are a heavy
 CC chain variable region and human-derived constant light and heavy chain
 CC regions. The variable regions have high affinity for TAG72. The constant
 CC regions reduce the side-effects when administered to human patients
 CC because they are of human origin.
 CC See also R04381-5 and R04386-8.
 SQ Sequence 116 AA;

Query Match 79.7% Score 59; DB 1; Length 116;
 Best Local Similarity 50.0%; Pred. No. 3.09e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 9 mllllwvsgt 18
 :||||| :|||
 QY 1 LLLLLWAAAA 10

RESULT 13
 ID R59509 standard; Protein; 133 AA.
 AC R59509;
 DT 31-DEC-1994 (first entry)
 DE Sequence of the light chain variable region of the mouse NA-8
 DE antibody, including the signal sequence and mature chain.
 KW Antibody NA-8; human CD18; complementarity determining region; CDR.
 OS Mus musculus.
 FH Key Location/Qualifiers
 FT protein 21..133
 FT /label= Mature light chain
 PN W09412214-A.
 PD 09-JUN-1994.
 PF 30-NOV-1993; U11611.
 PR 01-DEC-1992; US-983949.
 PA (PROT-) PROTEIN DESIGN LABS INC.
 PI Co MS, Landolfi NF;
 DR WPI: 94-199973/24.
 DR N-PSDB; Q66845.
 PT New humanised antibodies specific for CD18 - derived from new
 PT murine antibody NA-8, prevent binding of neutrophils to
 PT endothelial cells, useful for treating inflammation
 PS Disclosure; Fig 1A; 50pp; English.
 CC The mouse antibody NA-8 binds to human CD18. cDNAs for the heavy
 CC chain and light chain variable domain genes of NA-8 were cloned
 CC using anchored PCR. The cDNA variable domain sequences and the
 CC deduced AA sequences are shown in Q66845/R59509 and Q66846/R59510.
 SQ Sequence 133 AA;

Query Match 79.7% Score 59; DB 10; Length 133;
 Best Local Similarity 50.0%; Pred. No. 3.09e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 9 mllllwvsgt 18
 :||||| :|||
 QY 1 LLLLLWAAAA 10

RESULT 14
 ID R85909 standard; Protein; 133 AA.
 AC R85909;
 DT 02-JUL-1996 (first entry)
 DE Monoclonal antibody, CB-Hep.1, light chain variable domain.
 KW HBV; hepatitis B virus surface antigen; monoclonal; antibody;
 KW CB-Hep.1; immunopurification; detection; hybridoma; recombinant;
 KW Fv; variable; VH; HBSAg.
 OS Mus sp.
 FH Key Location/Qualifiers
 FT peptide 1..20
 FT /label= sig_peptide
 FT protein 21..133
 FT /label= mat_protein
 FT domain 21..43
 FT /label= FR1
 FT domain 44..60
 FT /label= CDR1

FT domain 61..75
 /label= FR2
 FT domain 76..82
 /label= CDR2
 FT domain 83..114
 /label= FR3
 FT domain 115..123
 /label= CDR3
 FT domain 124..133
 /label= FR4

PN EP-686696-A1.
 PD 13-DEC-1995.
 PF 09-JUN-1995; 201535.
 PR 09-JUN-1994; CU-000073.
 PA (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.
 PA (KAMB-) KAMBEEL R W.
 PI Ayala Avila MA, Canaan-Haden Frias LM; del Carmen Dominguez Horta MC;
 PI Fernandez de Cossio Dorta-Duque ME, Gavilondo Cowley JVG;
 DR WPI: 96-021914/03.
 DR N-PSDB; Q74067.
 PT New recombinant single chain Fv antibody fragment - useful for
 PT immunopurification and detection of HBSAg
 PS Claim 1; Page 19; 23pp; English.
 CC R85908 and R85909 are the heavy and light chain variable domains of
 CC the murine hybridoma-derived monoclonal antibody CB-Hep.1 Fv region.
 CC The heavy and light chain regions are used to produce a recombinant
 CC antibody fragment having specificity for the hepatitis B virus
 CC surface antigen (HBSAg). The recombinant antibody may be used for
 CC immunopurification of HBSAg, in immunoassays for the detection of
 CC HBSAg and for insolubilisation of heterologous fusion proteins
 CC expressed in E. coli aiding their purification.
 SQ Sequence 133 AA;

Query Match 79.7% Score 59; DB 16; Length 133;
 Best Local Similarity 50.0%; Pred. No. 3.09e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 9 mllllwvsgt 18
 :||||| :|||
 QY 1 LLLLLWAAAA 10

RESULT 15
 ID R30485 standard; Protein; 134 AA.
 AC R30485;
 DT 06-MAY-1993 (first entry)
 DE V kappa region of Ab to pre-S2 antigen of Hepatitis B virus .
 KW antibody; chimaeric; hepatitis B virus; pre-S2 surface antigen;
 KW treatment; prevention; enhanced immunogenicity; human C region; ss.
 OS Chimaeric; murine; human; mus musculus; homo sapiens.
 FH Key Location/Qualifiers
 FT region 1..16
 FT /label= leader
 PN EP-521348-A.
 PD 07-JAN-1993.
 PF 17-JUN-1992; 110262.
 PR 18-JUN-1991; KR-Q10040.
 PR 06-FEB-1992; KR-001692.
 PR 04-MAY-1992; KR-007593.
 PA (KOAD) KOREA ADV INST SCI & TECHN.
 PI Chung HK, Han MH, Hong HJ, Jin BR, Park SS, Ryu CJ;
 DR WPI: 93-001907/01.
 DR N-PSDB; Q34842.
 PT Chimaeric antibody specific for hepatitis B virus pre-S2 surface
 PT antigen - useful for treating and preventing hepatitis B virus
 PT infections
 PS Claim 3; Page 12; 23pp; English.
 CC This sequence represents that of the V kappa region of a chimaeric
 CC human/murine monoclonal antibody to the pre-S2 antigen. The pre-S2
 CC antigen is reported to have enhanced immunogenicity as compared with
 CC the currently used S antigen. Use of the human C region eliminates
 CC patient immune response to the murine C region which improves
 CC efficacy and reduces undesirable clinical side effects.

SQ Sequence 134 AA;
Query Match 79.7%; Score 59; DB 6; Length 134;
Best Local Similarity 50.0%; Pred. No. 3.09e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 9 mllllwvsgt 18
:|||||:::
Qy 1 LLLLLWAAAA 10

Search completed: Thu Jul 8 18:05:12 1999
Job time : 19 secs.

WQREH

(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:03:20 1999; MasPar time 5.19 Seconds
Tabular output not generated. 77.140 Million cell updates/sec

Title: >US-09-041-236-2
Description: (1-10) from US09041236.pep (1 of 45)
Perfect Score: 74
Sequence: 1 LLLLLWAAAA 10

Scoring table: PAM 150
Gap 11
Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4
Statistics: Mean 25.119; Variance 56.080; scale 0.448

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	67	90.5	354	2	I80170 class I histocompatib	3.58e+00
2	67	90.5	355	2	I37516 HLA-B alpha-chain - h	3.58e+00
3	67	90.5	359	1	HLHUB4 MHC class I histocomp	3.58e+00
4	67	90.5	362	2	I61861 MHC HLA-B*44.2 chain -	3.58e+00
5	67	90.5	362	1	HLHUB2 MHC class I histocomp	3.58e+00
6	67	90.5	362	2	I61905 MHC class I histocomp	3.58e+00
7	67	90.5	362	2	I61906 MHC class I histocomp	3.58e+00
8	67	90.5	362	2	I59655 lymphocyte antigen -	3.58e+00
9	67	90.5	362	2	I35997 MHC class I histocomp	3.58e+00
10	67	90.5	362	2	S52486 HLA-B protein alpha c	3.58e+00
11	67	90.5	362	2	I37515 MHC class I histocomp	3.58e+00
12	67	90.5	362	2	A45850 MHC class I histocomp	3.58e+00
13	67	90.5	362	2	I54289 MHC HLA-B*27d - human	3.58e+00
14	67	90.5	362	2	I61860 MHC HLA-B18 chain - h	3.58e+00
15	67	90.5	362	2	I68724 MHC class I histocomp	3.58e+00
16	67	90.5	362	2	I37485 human lymphocyte anti	3.58e+00
17	67	90.5	362	2	I56149 lymphocyte antigen -	3.58e+00
18	67	90.5	362	2	I54442 MHC class I histocomp	3.58e+00
19	67	90.5	362	2	S24415 class I histocompatib	3.58e+00
20	65	87.8	359	2	JCS382 hereditary hemochroma	6.51e+00
21	65	87.8	1838	1	CGHUIV collagen alpha 1(V) c	6.51e+00
22	64	86.5	362	2	I56130 HLA-B*5401 - human	8.76e+00
23	64	86.5	362	2	I72755 HLA-B*5602 - human	8.76e+00

24	64	86.5	362	2	I59633 MHC HLA-B transmembra	8.76e+00
25	64	86.5	362	2	I72752 HLA-B*5501 - human	8.76e+00
26	64	86.5	362	2	I72753 HLA-B*5502 - human	8.76e+00
27	64	86.5	362	2	I72754 HLA-B*5601 - human	8.76e+00
28	64	86.5	469	1	KCBOI interstitial collagen	8.76e+00
29	64	86.5	1843	2	S18803 collagen alpha 1(V) c	8.76e+00
30	62	83.8	468	1	KCRBI interstitial collagen	1.57e+01
31	61	82.4	406	1	REHUK rein (EC 3.4.23.15)	2.10e+01
32	61	82.4	1053	2	S44250 integrin alpha-5 chai	2.10e+01
33	61	82.4	1064	2	A41542 adenyate cyclase (EC	2.80e+01
34	60	81.1	202	2	G71295 conserved hypotheica	2.80e+01
35	60	81.1	362	2	I54298 gene HLA-B protein -	2.80e+01
36	60	81.1	362	2	A45880 MHC class I histocomp	2.80e+01
37	60	81.1	362	2	I37522 MHC class I histocomp	2.80e+01
38	60	81.1	362	2	S24435 class I histocompatib	2.80e+01
39	60	81.1	362	2	I37475 HLA-B35 variant - hum	2.80e+01
40	60	81.1	362	2	I61904 MHC class I histocomp	2.80e+01
41	60	81.1	362	2	I37120 MHC class I histocomp	2.80e+01
42	60	81.1	362	2	I56133 MHC class I protein -	2.80e+01
43	60	81.1	364	2	D35997 MHC class I histocomp	2.80e+01
44	60	81.1	445	2	S77010 nickel resistance pro	2.80e+01
45	60	81.1	605	2	A27274 ribophorin I precurso	2.80e+01

ALIGNMENTS

RESULT 1 I80170 #type fragment
ENTRY class I histocompatibility antigen - chimpanzee (fragment)
TITLE #formal_name Pan troglodytes #common_name chimpanzee
ORGANISM 24-May-1996 #sequence_revision 24-May-1996 #text_change
DATE 16-Feb-1997
ACCESSIONS I80170
REFERENCE I59308
#authors McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Watkins, D.I.
#journal Proc. Natl. Acad. Sci. U.S.A. (1994) 91:5893-5897
#title A uniquely high level of recombination at the HLA-B locus.
#cross-references EMBL:U05581; NID:g454779; PID:g454780
#accession I80170
#status preliminary; translated from GB/EMBL/DBBJ
#molecule_type mRNA
#residues 1-354 #label RPS
#cross-references EMBL:U05581; NID:g454779; PID:g454780
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 354 #checksum 5432

Query Match 90.5%; Score 67; DB 2: Length 354;
Best Local Similarity 80.0%; Pred No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1 LLLLLWGAVA 10
|||||:|
Qy 1 LLLLLWAAAA 10

RESULT 2 I37516 #type fragment
ENTRY HLA-B alpha-chain - human (fragment)
TITLE #formal_name Homo sapiens #common_name man
ORGANISM 04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change
DATE 16-Feb-1997
ACCESSIONS I37516
REFERENCE I37516
#authors Gauchat-Feiss, D.; Breur-Vriesendorp, B.S.; Rufer, N.; Jeannet, M.; Roosnek, E.; Tiercy, J.M.

#journal Tissue Antigens (1994) 44:261-264
#title Sequencing of a novel functional HLA-B*44 subtype differing in two residues in the alpha 2 domain.
#cross-references MIM:95176328
#accession I37516
#status preliminary; translated from GB/EMBL/DBBJ

```

##molecule_type mRNA
##residues 1-355 ##label RES
##cross-references EMBL:X75953; NID:g791007; PID:g791008
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 355 #checksum 1603

Query Match 90.5%; Score 67; DB 2; Length 355;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
QY 1 LLLLLWAAAA 10

RESULT 3
ENTRY HLHUB4 #type fragment
TITLE MHC class I histocompatibility antigen HLA-B44 alpha chain
ORGANISM precursor - human (fragment)
#formal_name Homo sapiens #common_name man
DATE 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change
20-Mar-1998
ACCESSIONS A25295
REFERENCE Kottmann, A.H.; Seemann, G.H.A.; Gueusov, H.D.; Roos, M.H.
#authors Immunogenetics (1986) 23:396-400
#journal DNA sequence of the coding region of the HLA-B44 gene.
#title #cross-references MUID:86249389
#accession A25295
#molecule_type mRNA
##residues 1-359 ##label KOT
##cross-references GB:M15470; NID:g187680; PID:g386883
GENETICS
#gene GDB:HLA-B
#map_position 6p21.3-6p21.3
#cross-references GB:120048; OMIM:142830
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
KEYWORDS duplication; glycoprotein; heterodimer; transmembrane
protein; transplantation antigen
FEATURE
1-21 #domain signal sequence (fragment) #status predicted
#label Sig\
22-359 #product class I histocompatibility antigen HLA-B44
alpha chain #status predicted #label MAT\
22-304 #domain extracellular #status predicted #label EXT\
22-111 #domain alpha-1 #label EX1\
112-203 #domain alpha-2 #label EX2\
217-282 #domain immunoglobulin homology #label IMM\
305-328 #domain transmembrane #status predicted #label TM\
329-359 #domain intracellular #status predicted #label INT\
107 #binding site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 359 #checksum 3910

Query Match 90.5%; Score 67; DB 1; Length 359;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 6 LLLLLWGAVA 15
QY 1 LLLLLWAAAA 10

RESULT 4
ENTRY I61861 #type complete
TITLE MHC HLA-B44.2 chain - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
16-Feb-1997
ACCESSIONS I61861
REFERENCE I36956

#authors Parham, P.; Lawlor, D.A.; Lomen, C.E.; Ennis, P.D.
#journal J. Immunol. (1989) 142:3937-3950
#title Diversity and diversification of HLA-A,B,C alleles.
#cross-references MUID:89235215
#accession I61861
##status preliminary; translated from GB/EMBL/DBDJ
##molecule_type DNA
##residues 1-362 ##label RES
##cross-references GB:M24038; NID:g187811; PID:g386900
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 362 #molecular-weight 40481 #checksum 618

Query Match 90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
QY 1 LLLLLWAAAA 10

RESULT 5
ENTRY HLHUB2 #type complete
TITLE MHC class I histocompatibility antigen HLA-B27 alpha chain
ORGANISM precursor - human
#formal_name Homo sapiens #common_name man
DATE 13-Aug-1986 #sequence_revision 28-Apr-1995 #text_change
05-Sep-1997
ACCESSIONS S07441; A25092; B25092; A94087; S34180; S44942; A90493;
B24741; I35965; S52291; A02184; A25128
REFERENCE S07441
#authors Weiss, E.H.; Kuon, W.; Doerner, C.; Lang, M.; Riethmueller,
G.
#journal Immunobiology (1985) 170:367-380
#title Organization, sequence and expression of the HLA-B27 gene: a
molecular approach to analyze HLA and disease associations.
#cross-references MUID:86138405
#accession S07441
##molecule_type DNA
##residues 1-362 ##label WEI
##cross-references EMBL:X03945
##note the authors translated the codon GAC for residue 61 as
Ala and the codon CAG for residue 111 as Glu
this allele is designated B*27052 (formerly 27W)
REFERENCE A91061
#authors Seemann, G.H.A.; Rein, R.S.; Brown, C.S.; Ploegh, H.L.
#journal EMBO J. (1986) 5:547-552
#title Gene conversion-like mechanisms may generate polymorphism in
human class I genes.
#cross-references MUID:86220133
#accession A25092
##molecule_type DNA
##residues 1-362 ##label SEE
##cross-references GB:X03665; NID:g32250; PID:g871297
##note this allele is designated B*27051 (formerly 27W)
#accession B25092
##molecule_type DNA
##residues 1-100,'N',102-103,'IA',106-362 ##label SE2
##cross-references GB:X03664; NID:g32236; PID:g871296
##note this allele is designated B*2702 (formerly 27K)
REFERENCE A94087
#authors Szoets, H.; Riethmueller, G.; Weiss, E.; Meo, T.
#journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:1428-1432
#title Complete sequence of HLA-B27 cDNA identified through the
characterization of structural markers unique to the HLA-A,
-B, and -C allelic series.
#cross-references MUID:86149317
#accession A94087
##molecule_type mRNA
##residues 25-205,'V',207-362 ##label SZO
##cross-references GB:M12678
##note this allele is designated B*27052 (formerly 27W)

```



```

REFERENCE S34180
#authors Vilches, C.
#submission submitted to the EMBL Data Library, June 1993
#accession S34180
#status preliminary
#molecule_type mRNA
#residues 1-100,'S',102-137,'D',139,'Y',141-175,'E',177-362
##label VIL
##cross-references EMBL:X73578
#note this allele is designated B*2706
REFERENCE S44942
#authors D'Amato, M.; Sorrentino, R.
#submission submitted to the EMBL Data Library, May 1994
#description Identification of a novel HLA-B*27 subtype by restriction
analysis of a cytotoxic gamma/delta T cell clone.
#accession S44942
#molecule_type mRNA
#residues 1-139,'H',141-362 ##label DAM
##cross-references EMBL:Z33453; NID:g486652; PID:g486653
REFERENCE A90493
#authors Ezquerro, A.; Bragado, R.; Vega, M.A.; Strominger, J.L.;
Woody, J.; Lopez de Castro, J.A.
#journal Biochemistry (1985) 24:1733-1741
#title Primary structure of papain-solubilized human
histocompatibility antigen HLA-B*27.
#cross-references MUID:85226361
#accession A90493
#molecule_type protein
#residues 25-265,'E',267-295 ##label EZQ
REFERENCE A94070
#authors Vega, M.A.; Ezquerro, A.; Rojo, S.; Aparicio, P.; Bragado,
R.; Lopez de Castro, J.A.
#journal Proc. Natl. Acad. Sci. U.S.A. (1985) 82:7394-7398
#title Structural analysis of an HLA-B*27 functional variant:
identification of residues that contribute to the
specificity of recognition by cytolytic T lymphocytes.
#cross-references MUID:86042671
#accession B24741
#molecule_type protein
#residues 86-100,'N',102-103,'IA',106-107,171-181 ##label VEG
REFERENCE I55965
#authors Coppin, H.L.; McDevitt, H.O.
#journal J. Immunol. (1986) 137:2168-2172
#title Absence of polymorphism between HLA-B*27 genomic exon
sequences isolated from normal donors and ankylosing
spondylitis patients.
#cross-references MUID:87009855
#accession I55965
#status preliminary; translated from GB/EMBL/DBBJ
#molecule_type DNA
#residues 25-298 ##label RES
##cross-references GB:M14013; NID:g187743; PID:g187744
REFERENCE S52291
#authors Blasczyk, R.; Weber, M.; Salama, A.
#submission submitted to the EMBL Data Library, January 1995
#accession S52291
#molecule_type DNA
#residues 116-192 ##label BLA
##cross-references EMBL:X83737
COMMENT This allele for HLA-B correlates with the development of ankylosing
spondylitis.
GENETICS
#gene GDB:HLA-B
#map_position 6p21.3-6p21.3
#classification #superfamily class I histocompatibility antigen;
immunoglobulin homology
transmembrane protein
KEYWORDS
FEATURE
1-24
25-362
#domain signal sequence #status predicted #label SIG\
#product MHC class I histocompatibility antigen HLA-B*40 alpha chain
#alpha chain #status predicted #label IMM\
#domain immunoglobulin homology #label IMM
#length 362 #molecular-weight 40430 #checksum 193
SUMMARY
Query Match 90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
|||||:|
Qy 1 LLLLLWAAAA 10

RESULT 7
ENTRY
TITLE
161906 #type complete
MHC class I histocompatibility antigen HLA-B*40 alpha chain
#precursor - human (isolate B*4004)
#formal_name Homo sapiens #common_name man
isolate B*4004
06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change
29-Aug-1997
ORGANISM
#variety
DATE
ACCESSIONS
161905
REFERENCE
I37120
#authors Belich, M.P.; Madrigal, J.A.; Hildebrand, W.H.; Zemmour, J.;
Williams, R.C.; Luz, R.; Petzi-Erler, M.L.; Parham, P.
#journal Nature (1992) 357:326-329
#title Unusual HLA-B alleles in two tribes of Brazilian Indians.
#cross-references MUID:92269955
#accession I61905
#status translated from GB/EMBL/DBBJ
#molecule_type mRNA
#residues 1-362 ##label RES
##cross-references GB:M84383; NID:g187715; PID:g187716
#experimental_source cell line GRC 138; isolate B*4003
GENETICS
#gene GDB:HLA-B
#map_position 6p21.3-6p21.3
#classification #superfamily class I histocompatibility antigen;
immunoglobulin homology
transmembrane protein
KEYWORDS
FEATURE
1-24
25-362
#domain signal sequence #status predicted #label SIG\
#product MHC class I histocompatibility antigen HLA-B*40 alpha chain
#alpha chain #status predicted #label IMM\
#domain immunoglobulin homology #label IMM
#length 362 #molecular-weight 40430 #checksum 193
SUMMARY
Query Match 90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
|||||:|
Qy 1 LLLLLWAAAA 10

RESULT 7
ENTRY
TITLE
161906 #type complete
MHC class I histocompatibility antigen HLA-B*40 alpha chain
#precursor - human (isolate B*4004)
#formal_name Homo sapiens #common_name man
isolate B*4004
06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change
29-Aug-1997
ORGANISM
#variety
DATE

```

```
ACCESSIONS      I61906
REFERENCE        I37120
#authors        Bellich, M.P.; Madrigal, J.A.; Hildebrand, W.H.; Zemmour, J.;
                  Williams, R.C.; Luz, R.; Petzl-Erler, M.L.; Parham, P.
#journal        Nature (1992) 357:326-329
#title          Unusual HLA-B alleles in two tribes of Brazilian Indians.
#cross-references MUID:92289955
#accession      I61906
#status         translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues       1-362 #label RES
#cross-references GB:M84384; NID:g187717; PID:g187718
#experimental_source cell line GRC 212; isolate B*4004
GENETICS
#gene           GDB:HLA-B
#map_position   6p21.3-6p21.3
#cross-references GDB:120048; OMIM:142830
CLASSIFICATION  #superfamily class I histocompatibility antigen;
                  transmembrane protein
KEYWORDS        1-24
                  #domain signal sequence #status predicted #label SIG\
25-362          #product MHC class I histocompatibility antigen HLA-B*40
                  alpha chain #status predicted #label MAT\
220-285         #domain immunoglobulin homology #label IMM
SUMMARY         #length 362 #molecular-weight 40600 #checksum 393
Query Match     90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 9 LLLLLWGAVA 18
   | | | | | : | |
QY 1 LLLLLWAAAA 10
RESULT 8
ENTRY      I59655 #type complete
TITLE      lymphocyte antigen - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
17-Mar-1999
ACCESSIONS I59655
REFERENCE   I59655
#authors    Adams, E.J.; Little, A.
#journal    Tissue Antigens (1995) 46:204-205
#title      Identification of a novel HLA-B allele (B*4008) in a patient
            with leukemia.
#cross-references MUID:96097410
#accession   I59655
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    1-362 #label RES
#cross-references GB:L41353; NID:g1041064; PID:g1041065
GENETICS
#gene        HLA-B*4008
#superfamily class I histocompatibility antigen;
            immunoglobulin homology
SUMMARY      #length 362 #molecular-weight 40550 #checksum 417
Query Match  90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 9 LLLLLWGAVA 18
   | | | | | : | |
QY 1 LLLLLWAAAA 10
RESULT 9
ENTRY      C35997 #type complete
TITLE      MHC class I histocompatibility antigen HLA-B*2706 alpha chain
            precursor - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change
29-Aug-1997
```

```
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            16-Nov-1990 #sequence_revision 13-Jan-1993 #text_change
24-Sep-1998
ACCESSIONS      C35997
REFERENCE        A35997
#authors        Emis, P.D.; Zemmour, J.; Salter, R.D.; Parham, P.
#journal        Proc. Natl. Acad. Sci. U.S.A. (1990) 87:2833-2837
#title          Rapid cloning of HLA-A,B cDNA by using the polymerase chain
                reaction: frequency and nature of errors produced in
                amplification.
#cross-references MUID:90207291
#accession      C35997
#status         preliminary
#molecule_type mRNA
#residues       1-362 #label ENN
#cross-references GB:M32320; NID:g187792; PID:g307224
GENETICS
#gene           GDB:HLA-B
#map_position   6p21.3-6p21.3
#cross-references GDB:120048; OMIM:142830
CLASSIFICATION  #superfamily class I histocompatibility antigen;
                  immunoglobulin homology
KEYWORDS        transmembrane protein
FEATURE         220-285
                  #domain immunoglobulin homology #label IMM
SUMMARY         #length 362 #molecular-weight 40456 #checksum 1360
Query Match     90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 9 LLLLLWGAVA 18
   | | | | | : | |
QY 1 LLLLLWAAAA 10
RESULT 10
ENTRY      S52486 #type complete
TITLE      HLA-B protein alpha chain - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       08-May-1995 #sequence_revision 21-Jul-1995 #text_change
08-Sep-1997
ACCESSIONS      S52486
REFERENCE        S52486
#authors        Herrero, M.J.
#submission     submitted to the EMBL Data Library, February 1995
#accession      S52486
#status         preliminary
#molecule_type mRNA
#residues       1-362 #label HER
#cross-references EMBL:X84725; NID:g683574; PID:g683575
CLASSIFICATION  #superfamily class I histocompatibility antigen;
                  immunoglobulin homology
FEATURE         220-285
                  #domain immunoglobulin homology #label IMM
SUMMARY         #length 362 #molecular-weight 40592 #checksum 651
Query Match     90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 9 LLLLLWGAVA 18
   | | | | | : | |
QY 1 LLLLLWAAAA 10
RESULT 11
ENTRY      I37515 #type complete
TITLE      MHC class I histocompatibility antigen HLA-B*2706 alpha chain
            precursor - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change
29-Aug-1997
```

```
ACCESSIONS I37515
REFERENCE
#authors Vilches, C.; de Pablo, R.; Kreisler, M.
#journal Immunogenetics (1994) 39:219
#title Nucleotide sequence of HLA-B*2706.
#cross-references MUID:9410284
#accession I37515
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-362 #label RES
#cross-references EMBL:X73578; NID:g975658; PID:g975659
GENETICS
#gene GDB:HLA-B
#cross-references GDB:120048; OMIM:142830
#map_position 6p21.3-6p21.3
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 362 #molecular-weight 40442 #checksum 8095
Query Match 90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 9 LLLLLWGAVA 18
| | | | | | | |
QY 1 LLLLLWAAAA 10
RESULT 12
ENTRY A45850 #type complete
TITLE MHC class I histocompatibility antigen HLA-B13.1 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change
17-Mar-1999
ACCESSIONS A45850
REFERENCE
#authors Kato, K.; Dupont, B.; Yang, S.Y.
#journal Immunogenetics (1989) 29:117-120
#title Localization of nucleotide sequence which determines
mongoloid subtype of HLA-B13.
#cross-references MUID:89122134
#accession A45850
#status preliminary
#molecule_type DNA
#residues 1-362 #label KAT
#cross-references GB:M24075; NID:g187703; PID:g386884
GENETICS
#gene GDB:HLA-B
#cross-references GDB:120048; OMIM:142830
#map_position 6p21.3-6p21.3
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
FEATURE
220-285 #domain immunoglobulin homology #label IMM
SUMMARY #length 362 #molecular-weight 40468 #checksum 9415
Query Match 90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 9 LLLLLWGAVA 18
| | | | | | | |
QY 1 LLLLLWAAAA 10
RESULT 13
ENTRY I54289 #type complete
TITLE MHC HLA-B*27d - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
29-Aug-1997
ACCESSIONS I54289
REFERENCE
```

```
#authors Choo, S.Y.; St. John, T.; Orr, H.T.; Hansen, J.A.
#journal Hum. Immunol. (1988) 21:209-219
#title Molecular analysis of the variant alloantigen HLA-B*27d
(HLA-B*2703) identifies a unique single amino acid
substitution.
#cross-references MUID:88227491
#accession I54289
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-362 #label RES
#cross-references GB:M54883; NID:g187663; PID:g187664
GENETICS
#gene GDB:HLA-B
#cross-references GDB:120048; OMIM:142830
#map_position 6p21.3-6p21.3
#introns 25/1; 115/1; 207/1; 299/1; 338/1; 349/1
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 362 #molecular-weight 40402 #checksum 6586
Query Match 90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 9 LLLLLWGAVA 18
| | | | | | | |
QY 1 LLLLLWAAAA 10
RESULT 14
ENTRY I61860 #type complete
TITLE MHC HLA-B18 chain - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
16-Feb-1997
ACCESSIONS I61860
REFERENCE
#authors Parham, P.; Lawlor, D.A.; Lomen, C.E.; Ennis, P.D.
#journal J. Immunol. (1989) 142:3937-3950
#title Diversity and diversification of HLA-A,B,C alleles.
#cross-references MUID:89235215
#accession I61860
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-362 #label RES
#cross-references GB:M24039; NID:g187809; PID:g386899
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 362 #molecular-weight 40275 #checksum 1354
Query Match 90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 9 LLLLLWGAVA 18
| | | | | | | |
QY 1 LLLLLWAAAA 10
RESULT 15
ENTRY I68724 #type complete
TITLE MHC class I histocompatibility antigen HLA-B*47 precursor -
human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change
14-Aug-1998
ACCESSIONS I68724
REFERENCE
#authors Zemmour, J.; Ennis, P.D.; Parham, P.; Dupont, B.
#journal Immunogenetics (1988) 27:281-287
#title Comparison of the structure of HLA-B*47 to HLA-B13 and its
relationship to 21-hydroxylase deficiency.
#cross-references MUID:88152906
```

```
#accession I68724
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-362 ##label ZEM
#cross-references GB:M19756; NID:G184171; PID:G386776
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 362 #molecular-weight 40571 #checksum 52

Query Match 90.5%; Score 67; DB 2; Length 362;
Best Local Similarity 80.0%; Pred No. 3.58e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
   |||||:|
Qy 1 LLLLLWAAAA 10
```

Search completed: Thu Jul 8 18:03:32 1999
Job time : 12 secs.

W P S R L

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:02:12 1999; Maspar time 3.46 seconds
Tabular output not generated. 81.739 Million cell updates/sec

Title: >US-09-041-236-2
Description: (1-10) from US09041236.pep (1 of 45)
Perfect Score: 74
Sequence: 1 LLLLLWAAA 10

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 25.801; Variance 51.563; scale 0.500

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Pred. No.	
Result No.	Score	Query Match	Length	ID	Description	
1	67	90.5	359	1	1B40_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
2	67	90.5	361	1	1B14_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
3	67	90.5	362	1	1B41_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
4	67	90.5	362	1	1B13_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
5	67	90.5	362	1	1B05_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
6	67	90.5	362	1	1B19_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
7	67	90.5	362	1	1B45_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
8	67	90.5	362	1	1B35_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
9	67	90.5	362	1	1B18_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
10	67	90.5	362	1	1B15_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
11	67	90.5	362	1	1B42_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
12	67	90.5	362	1	1B29_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
13	67	90.5	362	1	1B16_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
14	67	90.5	362	1	1B34_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
15	67	90.5	362	1	1B36_HUMAN	HLA CLASS I HISTOCOMPA 1.25e+00
16	65	87.8	359	1	HFE_MOUSE	HEREDITARY HAEMOCHROMA 2.42e+00
17	65	87.8	360	1	HFE_RAT	HEREDITARY HAEMOCHROMA 2.42e+00
18	65	87.8	1838	1	CA15_HUMAN	PROCOLLAGEN ALPHA 1(V) 2.42e+00
19	64	86.5	362	1	1B55_HUMAN	HLA CLASS I HISTOCOMPA 3.34e+00
20	64	86.5	362	1	1B56_HUMAN	HLA CLASS I HISTOCOMPA 3.34e+00
21	64	86.5	362	1	1B59_HUMAN	HLA CLASS I HISTOCOMPA 3.34e+00
22	64	86.5	362	1	1B58_HUMAN	HLA CLASS I HISTOCOMPA 3.34e+00
23	64	86.5	362	1	1B57_HUMAN	HLA CLASS I HISTOCOMPA 3.34e+00

ALIGNMENTS					PRT; 359 AA.	
RESULT	1	STANDARD;				
ID	1B40_HUMAN					
AC	P10320;					
DT	01-MAR-1989 (REL. 10, CREATED)					
DT	01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)					
DT	01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)					
DE	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-44(B-12) B*4401 ALPHA CHAIN					
DE	PRECURSOR (B44.1) (FRAGMENT).					
GN	HLA-B OR HLAB.					
OS	HOMO SAPIENS (HUMAN).					
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;					
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 86249389.					
RA	KOTTWANN A.H., SEEMANN G.H.A., GUESOW H.D., ROOS M.H.;					
RT	"DNA sequence of the coding region of the HLA-B44 gene.;"					
RL	IMMUNOGENETICS 23:396-400(1986).					
CC	-!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.					
CC	-!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; M15470; G386883; -					
DR	PIR; A25295; HLHUBA.					
DR	MIM; 142830; -					
DR	PROSITE; PS00290; IG_MHC; 1.					
DR	PFAM; PF00047; I; 1.					
DR	PFAM; PF00129; MHC_I; 1.					
DR	HSSP; P30491; IAIM.					
KW	MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.					
FT	NON_TER	1				
FT	SIGNAL	<1	21			
FT	CHAIN	22	359			
FT						
FT	DOMAIN	22	111			
FT	DOMAIN	112	203			
FT	DOMAIN	204	295			
FT	DOMAIN	296	305			
FT						

FT TRANSMEM 306 329
FT DOMAIN 330 359
FT CARBOHYD 107 107
FT DISULFID 122 185
FT DISULFID 224 280
SQ SEQUENCE 359 AA; 40040 MW; 7C0DB32E CRC32;

Query Match 90.5% Score 67; DB 1; Length 359;
Best Local Similarity 80.0%; Pred. No. 1.25e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 6 LLLLLGAVA 15
QY 1 LLLLLMAAA 10

RESULT 2
ID IB14-HUMAN STANDARD; PRT: 361 AA.
AC P03989;
DT 23-OCT-1986 (REL. 02, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 ALPHA CHAIN PRECURSOR.
GN HLA-B OR HLAB.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86138405.
RA WEISS E.H., KUON W., DOERNER C., LANG M., RIETHMUELLER G.;
RT "Organization, sequence and expression of the HLA-B27 gene: a
RT molecular approach to analyze HLA and disease associations.";
RL IMMUNOBIOLOGY 170:367-380(1985).
RN [2]
RP SEQUENCE OF 25-361 FROM N.A.
RX MEDLINE: 86149317.
RA SJOETS H., RIETHMUELLER G., WEISS E., MEO T.;
RT "Complete sequence of HLA-B27 cDNA identified through the
RT characterization of structural markers unique to the HLA-A, -B, and
RT -C allelic series";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:1428-1432(1986).
RN [3]
RP SEQUENCE OF 25-295.
RX MEDLINE: 85226361.
RA EQUEZERRA A., BRAGADO R., VEGA M.A., STROMINGER J.L., WOODY J.,
RA LOPEZ DE CASTRO J.A.;
RT "Primary structure of papain-solubilized human histocompatibility
RT antigen HLA-B27";
RL BIOCHEMISTRY 24:1733-1741(1985).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 25-300.
RX MEDLINE: 92405152.
RA MADDEN D.R., GORGA J.C., STROMINGER J.L., WILEY D.C.;
RT "The three-dimensional structure of HLA-B27 at 2.1-A resolution
RT suggests a general mechanism for tight peptide binding to MHC.";
RL CELL 70:1035-1048(1992).
RN [5]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE: 92018187.
RA MADDEN D.R., GORGA J.C., STROMINGER J.L., WILEY D.C.;
RT "The structure of HLA-B27 reveals nonamer self-peptides bound in an
RT extended conformation.";
RL NATURE 353:321-325(1991).
CC -I- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -I- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -I- DISEASE: THIS PROTEIN CORRELATES WITH THE DEVELOPMENT OF
CC ANKYLOSING SPONDYLITIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X03945; G32177; ALT_TERM.
CC PIR: A25128; HLHUB2.
CC PIR: S07441; S07441.
CC PDB: 1HSA; 15-OCT-92.
CC MIM: 142830.
CC PROSITE: PS00290; IG_MHC; 1.
CC PFAM: PF00047; IG_1.
CC PFAM: PF00129; MHC_I; 1.
KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL; 3D-STRUCTURE.
FT SIGNAL 1 24
FT CHAIN 25 361 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 25 114 ALPHA CHAIN B-27
FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.
FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.
FT TRANSMEM 309 332 CONNECTING PEPTIDE.
FT DOMAIN 333 361 CYTOPLASMIC TAIL.
FT CARBOHYD 110 110
FT DISULFID 125 188
FT DISULFID 227 283
FT CONFLICT 206 206
FT CONFLICT 266 266
FT CONFLICT 27 38
FT TURN 39 41
FT STRAND 42 52
FT TURN 53 54
FT STRAND 55 61
FT TURN 62 63
FT STRAND 70 71
FT HELIX 74 76
FT TURN 77 78
FT TURN 81 108
FT TURN 109 110
FT TURN 113 114
FT STRAND 118 127
FT TURN 129 130
FT STRAND 133 142
FT TURN 143 144
FT STRAND 145 150
FT TURN 152 153
FT STRAND 157 159
FT HELIX 162 173
FT TURN 174 175
FT HELIX 176 185
FT TURN 186 186
FT HELIX 187 198
FT TURN 199 199
FT HELIX 200 203
FT TURN 204 204
FT STRAND 207 207
FT STRAND 210 217
FT STRAND 222 233
FT STRAND 238 243
FT TURN 244 245
FT STRAND 246 247
FT HELIX 249 251
FT STRAND 253 254
FT STRAND 258 259
FT STRAND 265 274
FT TURN 275 276
FT HELIX 281 280
FT STRAND 281 286
FT TURN 288 289
FT STRAND 294 296
SQ SEQUENCE 361 AA; 40464 MW; 802130D5 CRC32;

RESULT	4	
ID	1B13_HUMAN	STANDARD;
AC	P30466;	PET; 362 AA.

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1].
RN SEQUENCE FROM N.A.
RP MEDLINE: 89235215.
RX

```
RA PARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;  
RT "Diversity and diversification of HLA-A,B,C alleles.";  
RL J. IMMUNOL. 142:3937-3950(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 88152906.  
RA ZEMMOUR J., ENNIS P.D., PARHAM P., DUPONT B.;  
RT "Comparison of the structure of HLA-B*47 and its  
RL relationship to 21-hydroxylase deficiency.";  
RN IMMUNOGENETICS 27:281-287(1988).  
[3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BLOOD;  
RX MEDLINE: 96053518.  
RA LIN L., TOKUNAGA K., NAKAJIMA F., ISHIKAWA Y., KASHIWASE K.,  
RA TANAKA H., KUMATA S., SIBELSEVA E., AKAZA T., TADOKORO K.,  
RA SHIBATA Y., CHANDANAYINGYONG D., JUJI T.;  
RT "Both HLA-B*1301 and B*1302 exist in Asian populations and are  
RL associated with different haplotypes.";  
RN HUM. IMMUNOL. 43:51-56(1995).  
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
CC THE IMMUNE SYSTEM.  
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: M24041; G386897; -  
CC EMBL: M19757; G386774; -  
CC EMBL: D50291; G1682953; -  
CC MIM: 142830; -  
CC PROSITE: PS00290; IG_MHC; 1.  
CC PFAM: PF00047; ig; 1.  
CC HSP: P30491; 1A1M.  
CC MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 24  
FT CHAIN 25 362  
FT DOMAIN 25 114  
FT DOMAIN 115 206  
FT DOMAIN 207 298  
FT DOMAIN 299 308  
FT TRANSMEM 309 332  
FT DOMAIN 333 362  
FT CARBOHYD 110 110  
FT DISULFID 125 188  
FT DISULFID 227 283  
FT CONFLICT 235 235  
SQ SEQUENCE 362 AA; 40474 MW; 28967875 CRC32;  
  
Query Match 90.5%; Score 67; DB 1; Length 362;  
Best Local Similarity 80.0%; Pred. No. 1.25e+00;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Db 9 LLLLLGAVA 18  
QY 1 LLLLLWAAAA 10  
| | | | | | | |  
RESULT 6  
ID 1B19_HUMAN STANDARD; PRT; 362 AA.  
AC Q08136;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2706 ALPHA CHAIN  
DE PRECURSOR.  
  
RA PARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;  
RT "Diversity and diversification of HLA-A,B,C alleles.";  
RL J. IMMUNOL. 142:3937-3950(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 88152906.  
RA ZEMMOUR J., ENNIS P.D., PARHAM P., DUPONT B.;  
RT "Comparison of the structure of HLA-B*47 and its  
RL relationship to 21-hydroxylase deficiency.";  
RN IMMUNOGENETICS 27:281-287(1988).  
[3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BLOOD;  
RX MEDLINE: 96053518.  
RA LIN L., TOKUNAGA K., NAKAJIMA F., ISHIKAWA Y., KASHIWASE K.,  
RA TANAKA H., KUMATA S., SIBELSEVA E., AKAZA T., TADOKORO K.,  
RA SHIBATA Y., CHANDANAYINGYONG D., JUJI T.;  
RT "Both HLA-B*1301 and B*1302 exist in Asian populations and are  
RL associated with different haplotypes.";  
RN HUM. IMMUNOL. 43:51-56(1995).  
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
CC THE IMMUNE SYSTEM.  
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: M24041; G386897; -  
CC EMBL: M19757; G386774; -  
CC EMBL: D50291; G1682953; -  
CC MIM: 142830; -  
CC PROSITE: PS00290; IG_MHC; 1.  
CC PFAM: PF00047; ig; 1.  
CC HSP: P30491; 1A1M.  
CC MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 24  
FT CHAIN 25 362  
FT DOMAIN 25 114  
FT DOMAIN 115 206  
FT DOMAIN 207 298  
FT DOMAIN 299 308  
FT TRANSMEM 309 332  
FT DOMAIN 333 362  
FT CARBOHYD 110 110  
FT DISULFID 125 188  
FT DISULFID 227 283  
FT CONFLICT 235 235  
SQ SEQUENCE 362 AA; 40474 MW; 28967875 CRC32;  
  
Query Match 90.5%; Score 67; DB 1; Length 362;  
Best Local Similarity 80.0%; Pred. No. 1.25e+00;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Db 9 LLLLLGAVA 18  
QY 1 LLLLLWAAAA 10  
| | | | | | | |  
RESULT 6  
ID 1B19_HUMAN STANDARD; PRT; 362 AA.  
AC Q08136;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2706 ALPHA CHAIN  
DE PRECURSOR.  
  
GN HLA-B OR HLAB.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94102824.  
RA VILCHES C., DE PABLO R., KREISLER M.;  
RT "Nucleotide sequence of HLA-B*2706.";  
RL IMMUNOGENETICS 39:219-219(1994).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96134006.  
RA RUDWALEIT M., BOWNES P., WORDSWORTH P.;  
RT "The nucleotide sequence of HLA-B*2704 reveals a new amino acid  
RT substitution in exon 4 which is also present in HLA-B*2706.";  
RL IMMUNOGENETICS 43:160-162(1996).  
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
CC THE IMMUNE SYSTEM.  
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: X73578; G975659; -  
CC EMBL: U35734; G1276641; -  
CC PROSITE: PS00290; IG_MHC; 1.  
CC PFAM: PF00047; ig; 1.  
CC HSP: P30491; 1A1M.  
CC MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 24  
FT CHAIN 25 362  
FT DOMAIN 25 114  
FT DOMAIN 115 206  
FT DOMAIN 207 298  
FT DOMAIN 299 308  
FT TRANSMEM 309 332  
FT DOMAIN 333 362  
FT CARBOHYD 110 110  
FT DISULFID 125 188  
FT DISULFID 227 283  
FT CONFLICT 235 235  
SQ SEQUENCE 362 AA; 40456 MW; 0D402027 CRC32;  
  
Query Match 90.5%; Score 67; DB 1; Length 362;  
Best Local Similarity 80.0%; Pred. No. 1.25e+00;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Db 9 LLLLLGAVA 18  
QY 1 LLLLLWAAAA 10  
| | | | | | | |  
RESULT 7  
ID 1B45_HUMAN STANDARD; PRT; 362 AA.  
AC P30485;  
DT 01-APR-1993 (REL. 25, CREATED)  
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-47 B*4701 ALPHA CHAIN  
DE PRECURSOR.  
GN HLA-B OR HLAB.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
```



```

RN SEQUENCE FROM N.A.
RX MEDLINE: 88152806.
RA ZEMMOUR J., ENNIS P.D., PARHAM P., DUPONT B.:
RT "Comparison of the structure of HLA-B*47 to HLA-B13 and its
RL relationship to 21-hydroxylase deficiency."
RL IMMUNOGENETICS 27:281-287(1988).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19756; G386776; -
CC MIM: 142830; -
CC PROSITE: PS00290; IG_MHC; 1.
CC PFAM: PF00047; ig; 1.
CC HSSP: P03989; 1HSA.
CC MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
CC SIGNAL 1 24
CC CHAIN 25 362
CC HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC ALPHA CHAIN BW-47 B*4701.
CC
CC DOMAIN 25 114
CC DOMAIN 115 206
CC DOMAIN 207 298
CC DOMAIN 299 309
CC TRANSMEM 310 333
CC DOMAIN 334 362
CC CARBOHYD 110 110
CC DISULFID 125 188
CC DISULFID 227 283
CC BY SIMILARITY.
CC SEQUENCE 362 AA; 40571 MW; 67734CIE CRC32;
CC
CC QUERY MATCH 90.5%; SCORE 67; DB 1; LENGTH 362;
CC BEST LOCAL SIMILARITY 80.0%; PRED. NO. 1.25e+00;
CC MATCHES 8; CONSERVATIVE 1; MISMATCHES 1; INDELS 0; GAPS 0;
CC
CC DB 9 LLLLLLGAVA 18
CC |||||:|
CC 1 LLLLLWAAAA 10
CC
CC RESULT 8
CC ID 1B35_HUMAN STANDARD; PRT; 362 AA.
CC AC P30477;
CC DT 01-APR-1993 (REL. 25, CREATED)
CC DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
CC DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
CC DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B*40 B*4003 ALPHA CHAIN
CC DE PRECURSOR (B40-G1).
CC GN HLA-B OR HLAB.
CC OS HOMO SAPIENS (HUMAN).
CC OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
CC OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE: 92269955.
CC RA BELICH M.P., MADRIGAL J.A., HILDEBRAND W.H., ZEMMOUR J.,
CC RA WILLIAMS R.C., LUZ R., PETZL-ERLER M.L., PARHAM P.:
CC "Unusual HLA-B alleles in two tribes of Brazilian Indians."
CC NATURE 357:326-329(1992).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M43484; G187718; -
CC MIM: 142830; -
CC PROSITE: PS00290; IG_MHC; 1.
CC PFAM: PF00047; ig; 1.
CC HSSP: P03989; 1HSA.
CC MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
CC SIGNAL 1 24
CC CHAIN 25 362
CC HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC ALPHA CHAIN B*40 B*4003.
CC
CC DOMAIN 25 114
CC DOMAIN 115 206
CC DOMAIN 207 298
CC DOMAIN 299 308
CC TRANSMEM 309 332
CC DOMAIN 333 362
CC CARBOHYD 110 110
CC DISULFID 125 188
CC DISULFID 227 283
CC BY SIMILARITY.
CC SEQUENCE 362 AA; 40600 MW; EIFE537C CRC32;
CC
CC QUERY MATCH 90.5%; SCORE 67; DB 1; LENGTH 362;
CC BEST LOCAL SIMILARITY 80.0%; PRED. NO. 1.25e+00;
CC MATCHES 8; CONSERVATIVE 1; MISMATCHES 1; INDELS 0; GAPS 0;
CC
CC DB 9 LLLLLLGAVA 18
CC |||||:|
CC 1 LLLLLWAAAA 10
CC
CC RESULT 9
CC ID 1B18_HUMAN STANDARD; PRT; 362 AA.
CC AC P10318;
CC DT 01-MAR-1989 (REL. 10, CREATED)
CC DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
CC DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
CC DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B*27 B*2705 ALPHA CHAIN
CC DE PRECURSOR (B-27W) (B27.1).
CC GN HLA-B OR HLAB.
CC OS HOMO SAPIENS (HUMAN).
CC OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
CC OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE: 86220133.
CC RA SEEMANN G.H.A., REIN R.S., BROWN C.S., PLOEGH H.L.:
CC "Gene conversion-like mechanisms may generate polymorphism in human
CC class I genes."
CC RL EMBO J. 5:547-552(1986).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE: 86138405.
CC RA WEISS E.H., KUON W., DOERNER C., LANG M., RIETHMUELLER G.:
CC "Organization, sequence and expression of the HLA-B27 gene: a
CC molecular approach to analyze HLA and disease associations."
CC RL IMMUNOBIOLOGY 170:367-380(1985).
CC RN [3]
CC RP 3D-STRUCTURE MODELLING OF 115-206.
CC RX MEDLINE: 95148615.
CC RA ROGNAN D., SCAPOZZA L., FOLKERS G., DASER A.:
CC "Rational design of nonnatural peptides as high-affinity ligands for
CC the HLA-B*2705 human leukocyte antigen."
CC PROC. NATL. ACAD. SCI. U.S.A. 92:753-757(1995).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO

```

```

CC      THE IMMUNE SYSTEM.
CC      -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC      MICROGLOBULIN).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: X03665; G871297; -
CC      EMBL: X03666; G871297; JOINED.
CC      EMBL: M12967; G187660; -
CC      PIR: A25097; HLHUBW.
CC      PDB: 1ROG; 30-SEP-94.
CC      PDB: 1ROH; 30-SEP-94.
CC      PDB: 1ROI; 30-SEP-94.
CC      PDB: 1ROJ; 30-SEP-94.
CC      PDB: 1ROK; 30-SEP-94.
CC      PDB: 1ROL; 30-SEP-94.
CC      MIM: 142830; -
CC      PROSITE: PS00290; IG_MHC; 1.
CC      PFAM: PF00047; Ig; 1.
CC      PFAM: PF00129; MHC_1; 1.
CC      MHC 1; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL; 3D-STRUCTURE.
CC      SIGNAL 1 24
CC      CHAIN 25 362
CC      DOMAIN 25 114
CC      DOMAIN 115 206
CC      DOMAIN 207 298
CC      DOMAIN 299 308
CC      TRANSMEM 309 332
CC      DOMAIN 333 362
CC      CARBOHYD 110 110
CC      DISULFID 125 188
CC      DISULFID 227 283
CC      SEQUENCE 362 AA; 40428 MW; 73243566 CRC32;
CC
CC      Query Match 90.5%; Score 67; DB 1; Length 362;
CC      Best Local Similarity 80.0%; Pred. No. 1.25e+00;
CC      Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
Db      9 LLLLLWGAVA 18
QY      1 LLLLLWAAAA 10
CC
RESULT 10
ID      1B15_HUMAN STANDARD; PRT; 362 AA.
AC      P10317;
DT      01-MAR-1989 (REL. 10, CREATED)
DT      01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT      01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE      HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2702 ALPHA CHAIN
DE      PRECURSOR (B*27K) (B27.2).
GN      HLA-B OR HLAB.
OS      HOMO SAPIENS (HUMAN).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC      PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
CC      [1]
RN      SEQUENCE FROM N.A.
RX      MEDLINE: 86220133.
RA      SEEMANN G.H.A., REIN R.S., BROWN C.S., PLOEGH H.L.;
RT      "Gene conversion-like mechanisms may generate polymorphism in human
RT      class I genes.";
RL      EMBO J. 5:547-552(1986).
RN      [2]
RN      SEQUENCE FROM N.A.
RA      PASHAM P., ARNETT K.L., ADAMS E.J.;
RT      SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

```

```

RN      SEQUENCE OF 86-107 AND 171-181.
RP      MEDLINE: 86042671.
RA      VEGA M.A., EZQUERRA A., ROJO S., APARICIO P., BRAGADO R.,
RA      LOPEZ DE CASTRO J.A.;
RT      "Structural analysis of an HLA-B27 functional variant: identification
RT      of residues that contribute to the specificity of recognition by
RT      cytolytic T lymphocytes.";
RL      CYCLOTIC T lymphocytes.";
RL      PROC. NATL. ACAD. SCI. U.S.A. 82:7394-7398(1985).
CC      -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC      THE IMMUNE SYSTEM.
CC      -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC      MICROGLOBULIN).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: X03664; G871296; -
CC      EMBL: X03667; G871296; JOINED.
CC      EMBL: L38504; G896271; -
CC      PIR: B25092; HLHUBK.
CC      MIM: 142830; -
CC      PROSITE: PS00290; IG_MHC; 1.
CC      PFAM: PF00047; Ig; 1.
CC      PFAM: PF00129; MHC_1; 1.
CC      HSP: P03989; IHSA.
CC      MHC 1; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
CC      SIGNAL 1 24
CC      CHAIN 25 362
CC      DOMAIN 25 114
CC      DOMAIN 115 206
CC      DOMAIN 207 298
CC      DOMAIN 299 308
CC      TRANSMEM 309 332
CC      DOMAIN 333 362
CC      CARBOHYD 110 110
CC      DISULFID 125 188
CC      DISULFID 227 283
CC      SEQUENCE 362 AA; 40397 MW; 9798F0BB CRC32;
CC
CC      Query Match 90.5%; Score 67; DB 1; Length 362;
CC      Best Local Similarity 80.0%; Pred. No. 1.25e+00;
CC      Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
Db      9 LLLLLWGAVA 18
QY      1 LLLLLWAAAA 10
CC
RESULT 11
ID      1B42_HUMAN STANDARD; PRT; 362 AA.
AC      P30482;
DT      01-APR-1993 (REL. 25, CREATED)
DT      01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT      15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE      HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-44 (B-12) B*4403 ALPHA CHAIN
DE      PRECURSOR.
GN      HLA-B OR HLAB.
OS      HOMO SAPIENS (HUMAN).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC      PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
CC      [1]
RN      SEQUENCE FROM N.A.
RX      MEDLINE: 91335451.
RA      FLEISCHHAUER K., KERNAN N.A., DUPONT B., YANG S.Y.;
RT      "The two major subtypes of HLA-B44 differ for a single amino acid in
RT      codon 156.";

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC -----
CC EMBL; M54883; G187664; -.
CC MIM; 142830; -.
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR PFAM; PF00047; IG; 1.
CC DR PFAM; PF00129; MHC_I; 1.
CC DR HSSP; P03989; ILSA.
CC KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
CC FT SIGNAL 1 24
CC FT CHAIN 25 362
CC FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC FT ALPHA CHAIN B-27 B*2703.
CC FT DOMAIN 25 114
CC FT DOMAIN 115 206
CC FT DOMAIN 207 298
CC FT DOMAIN 299 308
CC FT TRANSMEM 309 332
CC FT DOMAIN 333 362
CC FT CYTOPLASMIC TAIL.
CC FT CARBOHYD 110 110
CC FT BY SIMILARITY.
CC FT DISULFID 125 188
CC FT DISULFID 227 283
CC FT BY SIMILARITY.
CC SQ SEQUENCE 362 AA; 40402 MW; 7261C3AB CRC32;

Query Match 90.5%; Score 67; DB 1; Length 362;
Best Local Similarity 80.0%; Pred. No. 1.25e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
QY 1 LLLLLWAAAA 10

RESULT 14
ID 1B34_HUMAN STANDARD; PRT; 362 AA.
AC Q04826;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-40 B*4002 ALPHA CHAIN
DE PRECURSOR.
GN HLA-B OR HLAB.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93127148.
RA DOMENA J.D., JOHNSTON-DOM L., PARHAM P.;
RT "The B*4002 allele encodes the B61 antigen: B40* is identical to
RT B61.";
RL TISSUE ANTIGENS 40:254-256(1992).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
CC EMBL; M54883; G187664; -.
CC MIM; 142830; -.
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR PFAM; PF00047; IG; 1.
CC DR PFAM; PF00129; MHC_I; 1.
CC DR HSSP; P03989; ILSA.
CC KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
CC FT SIGNAL 1 24
CC FT CHAIN 25 362
CC FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC FT ALPHA CHAIN B-40 B*4002.
CC FT DOMAIN 25 114
CC FT DOMAIN 115 206
CC FT DOMAIN 207 298
CC FT DOMAIN 299 308
CC FT TRANSMEM 309 332
CC FT DOMAIN 333 362
CC FT CYTOPLASMIC TAIL.
CC FT CARBOHYD 110 110
CC FT BY SIMILARITY.
CC FT DISULFID 125 188
CC FT DISULFID 227 283
CC FT BY SIMILARITY.
CC SQ SEQUENCE 362 AA; 40505 MW; C2A00916 CRC32;

Query Match 90.5%; Score 67; DB 1; Length 362;
Best Local Similarity 80.0%; Pred. No. 1.25e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLWGAVA 18
QY 1 LLLLLWAAAA 10

RESULT 15
ID 1B36_HUMAN STANDARD; PRT; 362 AA.
AC P30478;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-40 B*4004 ALPHA CHAIN
DE PRECURSOR (B40-G2).
GN HLA-B OR HLAB.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92269955.
RA BELICH M.P., MADRIGAL J.A., HILDEBRAND W.H., ZEMMOUR J.,
RA WILLIAMS R.C., LUZ R., PETZL-ERLER M.L., PARHAM P.;
RT "Unusual HLA-B alleles in two tribes of Brazilian Indians.";
RL NATURE 357:326-329(1992).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

SQ SEQUENCE 362 AA; 40430 MW; 3797AB68 CRC32;

Query Match 90.5%; Score 67; DB 1; Length 362;
Best Local Similarity 80.0%; Pred. No. 1.25e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 LLLLLMGAVA 18

|||||

Qy 1 LLLLLWAAAA 10

Search completed: Thu Jul 8 18:02:21 1999
Job time : 9 secs.

THIS PAGE BLANK (USPTO)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:06:15 1999; MasPar time 8.04 Seconds
Tabular output not generated. 67.889 Million cell updates/sec

Title: >US-09-041-236-2
Description: (12-21) from US09041236.pep (2 of 45)
Perfect Score: 72
Sequence: 1 AOGHLRSGPR 10

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertebrate 14:sp.virus

Statistics: Mean 22.349; Variance 23.986; scale 0.932

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	72	100.0	666	4	SEMAPHORIN L.	2.10e-06
2	59	81.9	393	11	SEMAPHORIN L (FRAGMENT)	8.07e-03
3	54	75.0	58	11	CYTCHROME P450.2B15 (1.55e-01
4	54	75.0	310	14	HYPOTHETICAL PROTEIN (1.55e-01
5	54	75.0	327	14	THE FIRST ATG START CO	1.55e-01
6	54	75.0	383	14	HYPOTHETICAL BWR1 PRO	1.55e-01
7	54	75.0	437	14	THE FIRST ATG START CO	1.55e-01
8	52	72.2	57	11	CYTCHROME P450.2B13.	4.83e-01
9	52	72.2	491	11	CYTCHROME P450 (TESTO	4.83e-01
10	51	70.8	224	2	HYPOTHETICAL 23.6 KD P	8.44e-01
11	50	69.4	534	2	PUTATIVE PHOSPHOMANNOM	1.47e+00
12	48	66.7	122	2	HISTIDINE KINASE HOMOL	4.33e+00
13	48	66.7	183	3	40S RIBOSOMA PROTEIN.	4.33e+00
14	48	66.7	436	2	DEDUCED AMINOTRANSFERA	4.33e+00
15	48	66.7	571	2	ABSA1.	4.33e+00
16	48	66.7	591	10	CHLOROPLAST RNA POLYME	4.33e+00
17	47	65.3	353	10	GLYCERALDEHYDE-3-PHOSP	7.35e+00
18	47	65.3	424	4	8-OXOGUANINE DNA GLYCO	7.35e+00
19	47	65.3	1083	2	ACRIFLAVIN RESISTANCE	7.35e+00
20	47	65.3	1140	11	COLLAGEN (FRAGMENT).	7.35e+00

21	47	65.3	1360	5	Q20786	7.35e+00
22	47	65.3	1379	5	Q17353	7.35e+00
23	47	65.3	1527	11	Q82002	7.35e+00
24	47	65.3	1774	11	Q62001	7.35e+00
25	47	65.3	2037	5	Q22511	7.35e+00
26	46	63.9	108	14	Q67695	1.24e+01
27	46	63.9	118	4	Q13541	1.24e+01
28	46	63.9	534	2	Q69553	1.24e+01
29	46	63.9	534	13	Q90867	1.24e+01
30	46	63.9	743	4	Q43373	1.24e+01
31	46	63.9	1212	4	Q60305	1.24e+01
32	46	63.9	2911	5	Q33442	1.24e+01
33	45	62.5	72	14	Q68557	2.07e+01
34	45	62.5	349	10	Q81414	2.07e+01
35	45	62.5	388	2	Q32848	2.07e+01
36	45	62.5	491	11	Q84463	2.07e+01
37	45	62.5	594	11	Q88290	2.07e+01
38	45	62.5	598	3	Q60162	2.07e+01
39	45	62.5	663	5	Q93518	2.07e+01
40	45	62.5	682	5	Q17583	2.07e+01
41	45	62.5	800	5	Q93903	2.07e+01
42	45	62.5	1519	2	Q48237	2.07e+01
43	45	62.5	1894	10	Q64795	2.07e+01
44	44	61.1	422	14	Q99305	3.42e+01
45	44	61.1	1571	4	Q43161	3.42e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	666	AA.
ID	075326				
AC	075326;	1998 (TREMREL. 08, CREATED)			
DT	01-NOV-1998	(TREMREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMREL. 08, LAST SEQUENCE UPDATE)			
DE	SEMAPHORIN L.				
GN	SEMAL.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;				
OC	CATARRHINI; HOMINIDAE; HOMO.				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98389619.				
RA	LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;				
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA				
RT	viruses.";				
RL	GENOMICS 51:340-350(1998).				
DR	EMBL; AF030698; G3523115; --				
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;				

Query Match 100.0%; Score 72; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 2.10e-06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	44	AQHLRSGPR 53	
Qy	12	AQHLRSGPR 21	
RESULT	2	PRELIMINARY;	PRT; 393
ID	08371		
AC	08371;		
DT	01-NOV-1998	(TREMREL. 08, CREATED)	
DT	01-NOV-1998	(TREMREL. 08, LAST SEQUENCE UPDATE)	
DE	SEMAPHORIN L (FRAGMENT).		
GN	SEMAL.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 98389619.		

RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL: AF030699; G3523117; -.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;

Query Match 81.9%; Score 59; DB 11; Length 393;
Best Local Similarity 90.0%; Pred. No. 8.07e-03;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 44 AQGHSRSGPR 53
QY 12 AQGHLRSRSGPR 21
|||||
|||||

RESULT 3
ID Q63140 PRELIMINARY; PRT; 58 AA.
AC Q63140;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 2B15 (FRAGMENT).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RA NAKAYAMA K.;
RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE: 94124025.
RA NAKAYAMA K., SUWA Y., MIYUKAMI Y., SOGAWA K., FUJII-KURIYAMA Y.;
RT "Cloning and sequencing of a novel rat cytochrome p450 2B-encoding
gene."
RL GENE 136:333-336(1993).
DR EMBL: D17343; D1004681; -.
FT NON_TER 58 58
SQ SEQUENCE 58 AA; 6522 MW; 60E1D1D9 CRC32;

Query Match 75.0%; Score 54; DB 11; Length 58;
Best Local Similarity 60.0%; Pred. No. 1.55e-01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 27 THGHLPPGPR 36
QY 12 AQGHLRSRSGPR 21
:|||||
:

RESULT 4
ID Q69130 PRELIMINARY; PRT; 310 AA.
AC Q69130;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN (FRAGMENT).
OS HUMAN HERPESVIRUS TYPE 4.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-P3HR-1 (HR-1);
RX MEDLINE: 87198876.
RA JENSON H.B., FARRELL P.J., MILLER G.;
RT "Sequences of the Epstein-Barr Virus (EBV) large internal repeat form
the center of a 16-kilobase-pair palindrome of EBV (P3HR-1)
heterogeneous DNA."
RL J. VIROL. 61:1495-1506(1987).
DR EMBL: M15973; G808659; -.

KW HYPOTHETICAL PROTEIN.
FT NON_TER 1 1
SQ SEQUENCE 310 AA; 32159 MW; B3D062E2 CRC32;

Query Match 75.0%; Score 54; DB 14; Length 310;
Best Local Similarity 70.0%; Pred. No. 1.55e-01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 232 GOGHLRVPPR 241
QY 12 AQGHLRSRSGPR 21
:|||||
:

RESULT 5
ID Q69145 PRELIMINARY; PRT; 327 AA.
AC Q69145;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE THE FIRST ATG START CODON IS THE AA BEFORE THE STOP CODON IN ORF1
(FRAGMENT).
OS HUMAN HERPESVIRUS TYPE 4.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-P3HR-1 (HR-1);
RX MEDLINE: 87198876.
RA JENSON H.B., FARRELL P.J., MILLER G.;
RT "Sequences of the Epstein-Barr Virus (EBV) large internal repeat form
the center of a 16-kilobase-pair palindrome of EBV (P3HR-1)
heterogeneous DNA."
RL J. VIROL. 61:1495-1506(1987).
DR EMBL: M15972; G808667; -.
FT NON_TER 1 1
SQ SEQUENCE 327 AA; 33982 MW; 1321F99A CRC32;

Query Match 75.0%; Score 54; DB 14; Length 327;
Best Local Similarity 70.0%; Pred. No. 1.55e-01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 91 GOGHLRVPPR 100
QY 12 AQGHLRSRSGPR 21
:|||||
:

RESULT 6
ID Q04397 PRELIMINARY; PRT; 383 AA.
AC Q04397;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL BRF1 PROTEIN.
GN BRF1 OR BCRF2.
OS EPSTEIN-BARR VIRUS (STRAIN B95-8).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE; LYMPHOCRYPTOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84270667.
RA BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
RA GIBSON T.J., HATFULL G., HUDSON G.S., SATCHWELL S.C., SEGUIN C.,
RA TUFFNELL P.S., BARRELL B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus
genome."
RL NATURE 310:207-211(1984).
RN [2]
RP SEQUENCE OF 88-145 FROM N.A.
RC STRAIN-B95-8;
RX MEDLINE: 89296424.
RA WALLS D., CANNON F.;
RT "The expression of novel antigens from the Epstein-Barr virus large


```
RT internal repeat."
RL EMBO J. 7:1191-1196(1988).
DR EMBL: V01555; E25067; -
DR EMBL: V01555; E25068; -
DR EMBL: V01555; E25069; -
DR EMBL: V01555; E25070; -
DR EMBL: V01555; E25071; -
DR EMBL: V01555; E25066; -
DR EMBL: V01555; E25072; -
DR EMBL: X07816; G93868; -
DR EMBL: V01555; E25077; -
DR EMBL: V01555; E25078; -
DR EMBL: V01555; E25079; -
DR EMBL: V01555; E25080; -
DR EMBL: V01555; E25081; -
KW HYPOTHETICAL PROTEIN
SQ SEQUENCE 383 AA; 39866 MW; F63240D6 CRC32;

Query Match 75.0%; Score 54; DB 14; Length 383;
Best Local Similarity 70.0%; Pred. No. 1.55e-01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 147 GQHRLVPPR 156
QY 12 AQHLRSGPR 21

RESULT 7
ID Q69146 PRELIMINARY; PRT; 437 AA.
AC Q69146;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE THE FIRST ATG START CODON IS THE AA BEFORE THE STOP CODON IN ORF2
DE (FRAGMENT).
OS HUMAN HERPESVIRUS TYPE 4.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-P3HR-1 (HR-1);
RX MEDLINE: 87198876.
RA JENSON H.B., FARRELL P.J., MILLER G.;
RT "Sequences of the Epstein-Barr Virus (EBV) large internal repeat form
RT the center of a 16-kilobase-pair palindrome of EBV (P3HR-1)
RT heterogeneous DNA."
RL J. VIROL. 61:1495-1506(1987).
DR EMBL: M15972; G808668; -
DR NON_TER 1
FT SEQUENCE 437 AA; 45415 MW; D54090A1 CRC32;

Query Match 75.0%; Score 54; DB 14; Length 437;
Best Local Similarity 70.0%; Pred. No. 1.55e-01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 201 GQHRLVPPR 210
QY 12 AQHLRSGPR 21

RESULT 8
ID Q61460 PRELIMINARY; PRT; 57 AA.
AC Q61460;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450, 2B13, PHENOBARBITOL INDUCIBLE, TYPE C
DE (TESTOSTERONE 16A-HYDROXYLASE TYPE C) (FRAGMENT).
GN CYP2B13 OR 16A0H-C.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE: 91146586.
RA LAKSO M., MASAKI R., NOSHIRO M., NEGISHI M.;
RT "Structures and characterization of sex-specific mouse cytochrome
RT P-450 genes as members within a large family. Duplication boundary
RT and evolution."
RL EUR. J. BIOCHEM. 195:477-486(1991).
DR EMBL: M60359; G192890; -
DR MGD: MGI:88599; CYP2B13.
KW TESTOSTERONE.
FT NON_TER 57
SQ SEQUENCE 57 AA; 6311 MW; 4E241316 CRC32;

Query Match 72.2%; Score 52; DB 11; Length 57;
Best Local Similarity 66.7%; Pred. No. 4.83e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 27 HGHLPFGPR 35
QY 13 QGHLRSGPR 21

RESULT 9
ID Q64460 PRELIMINARY; PRT; 491 AA.
AC Q64460;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 (TESTOSTERONE 16-ALPHA HYDROXYLASE TYPE B)
DE (EC 1.14.14.1).
GN 16A0H-B.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91146586.
RA LAKSO M., MASAKI R., NOSHIRO M., NEGISHI M.;
RT "Structures and characterization of sex-specific mouse cytochrome
RT P-450 genes as members within a large family. Duplication boundary
RT and evolution."
RL EUR. J. BIOCHEM. 195:477-486(1991).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: M60358; G192863; -
DR EMBL: M60352; G192863; JOINED.
DR EMBL: M60353; G192863; JOINED.
DR EMBL: M60355; G192863; JOINED.
DR EMBL: M60354; G192863; JOINED.
DR EMBL: M60357; G192863; JOINED.
DR EMBL: M60356; G192863; JOINED.
DR MGD: MGI:88599; CYP2B13.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR PFAM: PF00067; P450; 1.
KW TESTOSTERONE; OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT;
KW MEMBRANE; HEME.
FT BINDING 436 436 HEME (BY SIMILARITY).
SQ SEQUENCE 491 AA; 55759 MW; AC36D7D0 CRC32;

Query Match 72.2%; Score 52; DB 11; Length 491;
Best Local Similarity 66.7%; Pred. No. 4.83e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 27 HGHLPFGPR 35
QY 13 QGHLRSGPR 21

RESULT 10
ID P96938 PRELIMINARY; PRT; 224 AA.
AC P96938;
DT 01-NAY-1997 (TREMBLREL. 03, CREATED)
DT 01-NAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
```

DE HYPOTHETICAL 23.6 KD PROTEIN.
GN MTCY20H10.30.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
RN ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA OLIVER K., HARRIS D.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT leprae";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
DR EMBL; Z92772; E306578; -;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 224 AA; 23616 MW; E5687B81 CRC32;

Query Match 70.8%; Score 51; DB 2; Length 224;
Best Local Similarity 60.0%; Pred. No. 8.44e-01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 41 AEARLRDGP 50
QY 12 AOGHLRSGP 21
I:::|||||

RESULT 11
ID O53360 PRELIMINARY; PRT; 534 AA.
AC O53360.
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE PHOSPHOMUTASE.
GN MT016.07.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
RN ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA OLIVER K., HARRIS D.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT leprae";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
DR EMBL; AL021841; E1251139; -;
DR PROSITE; PS00710; PGM_PMM; 1.

SQ SEQUENCE 534 AA; 56195 MW; 4FDE95B4 CRC32;
Query Match 69.4%; Score 50; DB 2; Length 534;
Best Local Similarity 75.0%; Pred. No. 1.47e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 43 RGHRLRGGP 50
QY 13 QGHRLRSGP 20
:|||||

RESULT 12
ID P72390 PRELIMINARY; PRT; 122 AA.
AC P72390;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HISTIDINE KINASE HOMOLOG (ABSAL) AND RESPONSE REGULATOR HOMOLOG
DE (ABSAL) GENES, COMPLETE CDS (FRAGMENT).
GN D9.
OS STREPTOMYCES COELICOLOR.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
RN ACTINOMYCETALES; STREPTOMYCINEAE; STREPTOMYCETACEAE; STREPTOMYCES.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-J1501;
RX MEDLINE; 96236039.
RA BRIAN P., RIGGLE P.J., SANTOS R.A., CHAMPNESS W.C.;
RT "Global negative regulation of Streptomyces coelicolor antibiotic
RT synthesis mediated by an absA-encoded putative signal transduction
RT system";
RL J. BACTERIOL. 178:3221-3231(1996).
DR EMBL; U51332; G1542978; -;
FT NON_TER 1
SQ SEQUENCE 122 AA; 13317 MW; 0B1392A1 CRC32;

Query Match 66.7%; Score 48; DB 2; Length 122;
Best Local Similarity 70.0%; Pred. No. 4.33e+00;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 111 AAGPLRQGP 120
QY 12 AOGHLRSGP 21
I:::|||||

RESULT 13
ID O13835 PRELIMINARY; PRT; 183 AA.
AC O13835;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 40S RIBOSOMA PROTEIN.
GN SPAC19D5.05C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
RN SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RA DEVLIN K., CHURCHER C.M.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RA WOOD V., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Z99531; E351302; -;
DR PRAM; PF00163; S4; 1.
SQ SEQUENCE 183 AA; 21777 MW; 3A501D45 CRC32;

Query Match 66.7%; Score 48; DB 3; Length 183;
Best Local Similarity 66.7%; Pred. No. 4.33e+00;

Job time : 16 secs.

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 133 OGHVRVGP 141

QY 13 OGHLSGPR 21

RESULT 14
ID O66156 PRELIMINARY; PRT; 436 AA.
AC O66156;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE DEDUCED AMINOTRANSFERASE.
GN KASC.
OS STREPTOMYCES KASUGAENSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M338-M1.
RX MEDLINE; 98250958.
RA IKENO S., TSUJI T., HIGASHIDE K., KINOSHITA N., HAMADA M., HORI M.;
RT "A 7.6kb DNA region from Streptomyces kasugaensis M338-M1 includes
RT some genes responsible for kasugamycin biosynthesis.";
RL J. ANTIBIOT. 51:341-352(1998).
DR EMBL; AB005901; D1026586;
DR PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
DR TRANSFERASE; AMINOTRANSFERASE.
SQ SEQUENCE 436 AA; 45394 MW; EB2E4B3E CRC32;

Query Match 66.7%; Score 48; DB 2; Length 436;

Best Local Similarity 75.0%; Pred. No. 4.33e+00;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 290 ARLRSGPR 297

QY 14 GHLRSGPR 21

RESULT 15
ID Q53893 PRELIMINARY; PRT; 571 AA.
AC Q53893;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE ABSA1.
DE ABSA1.
GN ABSA1.
OS STREPTOMYCES COELICOLOR.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J1501.
RX MEDLINE; 96236039.
RA BRIAN P., RIGGLE P.J., SANTOS R.A., CHAMPNESS W.C.;
RT "Global negative regulation of Streptomyces coelicolor antibiotic
RT synthesis mediated by an absA-encoded putative signal transduction
RT system.";
RL J. BACTERIOL. 178:3221-3231(1996).
DR EMBL; U51332; G1293659;
SQ SEQUENCE 571 AA; 61142 MW; E3A8D137 CRC32;

Query Match 66.7%; Score 48; DB 2; Length 571;

Best Local Similarity 70.0%; Pred. No. 4.33e+00;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 369 AGTLRTGPR 378

QY 12 AQGLRSGPR 21

Search completed: Thu Jul 8 18:06:31 1999

THIS PAGE BLANK (USPTO)

W P E R E L I
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:07:22 1999; Maspar time 9.53 Seconds
Tabular output not generated. 22.324 Million cell updates/sec

Title: >US-09-041-236-2
Description: (12-21) from US09041236.pep (2 of 45)
Perfect Score: 72
Sequence: 1 AOGHLRSQPR 10

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 16.539; Variance 44.296; scale 0.373

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	51	70.8	492	1	P94263 Bovine adrenal gland	3.18e+01
2	47	65.3	43	18	R98295 NP5 fusion protein.	9.54e+01
3	47	65.3	93	18	R96579 Monoclonal antibody M	9.54e+01
4	46	63.9	3010	16	R95022 Hepatitis GB virus (H	1.25e+02
5	45	62.5	300	22	W18642 Fragmented human amyl	1.63e+02
6	44	61.1	351	30	W40056 P300/CBP-associated t	2.13e+02
7	44	61.1	529	28	W44155 Human neuronal nicoti	2.13e+02
8	44	61.1	529	20	W09021 Neuronal nicotinic ac	2.13e+02
9	44	61.1	529	14	R73966 Alpha 2 subunit of a	2.13e+02
10	44	61.1	813	30	W40059 Mouse P300/CBP-associ	2.13e+02
11	44	61.1	832	30	W40052 Human P300/CBP-associ	2.13e+02
12	44	61.1	1464	38	W79294 An antigen from dermo	2.13e+02
13	43	59.7	36	1	P91333 Amino acid sequence o	2.77e+02
14	43	59.7	36	1	P91332 Amino acid sequence o	2.77e+02
15	43	59.7	63	10	R53354 sCEC-Shiva I fusion p	2.77e+02
16	43	59.7	93	25	W22078 Rat amylin.	2.77e+02

17	43	59.7	265	3	R12844 HTLV-1 protein expres	2.77e+02
18	43	59.7	332	35	W76189 Actinoplanes sp. acar	2.77e+02
19	43	59.7	492	2	P70387 gag nucleoprotein of	2.77e+02
20	43	59.7	629	15	R88466 Drosophila scavenger	2.77e+02
21	43	59.7	1093	9	R56978 Human myotonic dystro	2.77e+02
22	43	59.7	1487	36	W76720 Thermococcus 9N2 muta	2.77e+02
23	43	59.7	1487	36	W76719 Thermococcus 9N2 beta	2.77e+02
24	42	58.3	197	39	W88494 Human stomach cancer	3.60e+02
25	42	58.3	299	39	W80985 B cell receptor assoc	3.60e+02
26	42	58.3	299	39	W80984 Amino acid sequence o	3.60e+02
27	42	58.3	299	34	W61629 Clone HMWGS46 of Proh	3.60e+02
28	42	58.3	376	37	W81747 M. tuberculosis immun	3.60e+02
29	42	58.3	376	34	W64377 Mycobacterium tubercu	3.60e+02
30	42	58.3	377	34	W70902 Human cytochrome P-45	3.60e+02
31	42	58.3	434	18	R96420 Peptide fragment of N	3.60e+02
32	42	58.3	454	38	W85046 Gamma subunit of a DN	3.60e+02
33	42	58.3	464	38	W85045 Gamma subunit of a DN	3.60e+02
34	42	58.3	529	38	W85044 Tau subunit of a DNA	3.60e+02
35	42	58.3	622	29	W41248 Thermococcus amidase.	3.60e+02
36	42	58.3	729	8	R41753 Catalase-R.	3.60e+02
37	42	58.3	730	8	R41543 Aspergillus niger cat	3.60e+02
38	42	58.3	1089	38	W82396 Human UDP protein #2.	3.60e+02
39	42	58.3	1931	5	R27649 Human calcium channel	3.60e+02
40	42	58.3	2237	6	R33550 Sequence of the alpha	3.60e+02
41	42	58.3	2237	14	R71006 Human neuronal calciu	3.60e+02
42	42	58.3	2337	32	W37878 Human calcium channel	3.60e+02
43	42	58.3	2339	14	R71005 Human neuronal calciu	3.60e+02
44	42	58.3	2339	6	R33549 Sequence of the alpha	3.60e+02
45	42	58.3	2339	33	W63141 Human calcium channel	3.60e+02

ALIGNMENTS

RESULT 1

ID P94263 standard; protein; 492 AA.
AC P94263:
DT 19-JUN-1990 (first entry)
DE Bovine adrenal gland adrenodoxin reduction enzyme.
KW Bovine adrenal gland adrenodoxin reduction enzyme;
KW electron transport system.
OS Bos taurus.
PN J01010989-A.
PF 13-JAN-1989.
PD 30-JUN-1987; 165084.
PR 30-JUN-1987; JP-165084.
PA (KATA/) Katanaka Y.
DR N-PSDB; N91045.
PT Gene encoding bovine adrenal gland adrenodoxin reduction enzyme - used in
PT prodn. of enzymes composing electron transport system of bovine adrenal
PT gland mitochondria.
PS Disclosure; 2.; 7pp; Japanese.
CC The adrenodoxin reduction enzyme produces all the enzymes composing the
CC electron transport system of bovine adrenal gland mitochondria.
SQ Sequence 492 AA;

Query Match 70.8%; Score 51; DB 1; Length 492;
Best Local Similarity 87.5%; Pred. No. 3.18e+01;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 429 ghlpssgr 436
||| ||||
Qy 14 GHLRSQPR 21

RESULT 2

ID R98295 standard; Protein; 43 AA.
AC R98295:
DT 09-OCT-1996 (first entry)
DE NP5 fusion protein.
KW Breast cancer; MCF7; antigen; tumour marker; immunoassay;
KW monoclonal antibody; mammary tumor; metastasis; diagnosis.
OS Homo sapiens.

```

FH Key          Location/Qualifiers
FT region       18..22 "epitope that binds MAb Mc5"
PT
PN US5532135-A.
PD 02-JUL-1996.
PF 02-FEB-1990; 473673.
PR 02-FEB-1990; US-473673.
PR 08-APR-1993; US-046103.
PA (CANC-) CANCER RES FUND CONTRA COSTA.
PI Ceriani RL, Larocca DJ, Peterson JA;
DR WPI: 96-32111/32.
DR N-PSDB: T34468.
PT Competitive immunoassay for peptide antigen, esp. tumour marker -
PT using immobilised fusion protein as competitive reagent.
PS Example 6; Column 23-24; 17pp; English.
CC NP5 fusion protein (R98295) is produced by NP5 cDNA (T34468)
CC inserted into lambda/gtl1 and expressed in bacteria infected with
CC the phage carrying the cDNA immediately after the beta-galactosidase
CC gene. It contains an epitope (R98296) for antibody Mc5, whose native
CC antigen is human breast mucin NPgp. NP5 does not bind to other
CC anti-NPgp MABs. The results illustrate the use of MABs to select a
CC DNA fragment from a cDNA library that encodes a portion of a non-
CC natural fusion protein that is produced in bacteria as a result of
CC being inserted after an auxiliary gene (i.e. beta-galactosidase).
CC The fusion protein can be incorporated into a kit and/or immunoassay
CC to detect neoplastic peptides from mammary tumour or metastasis.
SQ Sequence 43 AA;

Query Match 65.3%; Score 47; DB 18; Length 43;
Best Local Similarity 75.0%; Pred. No. 9.54e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 14 qgdrlrpgp 21
QY 13 QGHLRSGP 20

RESULT 3
ID R96579 standard; Protein: 93 AA.
AC R96579;
DE Monoclonal antibody Mc5-binding portion of a beta-galactosidase/
DE breast cancer antigen fusion protein, NP5.
DE Fusion protein; solid-phase immunoassay; analyte; antibody;
KW specificity; detection; immobilise; breast cancer; MCF-7.
OS Synthetic.
FH Key          Location/Qualifiers
FT misc_difference 1..93
FT "note" "X residues correspond to stop codons
FT in file T28174"
PN US5514558-A.
PD 07-MAY-1996.
PF 02-FEB-1990; 473673.
PR 02-FEB-1990; US-473673.
PR 08-APR-1993; US-046103.
PR 30-SEP-1993; US-129540.
PA (CANC-) CANCER RES FUND CONTRA COSTA.
PI Ceriani RL, Larocca DJ, Peterson JA;
DR WPI: 96-238766/24.
DR N-PSDB: T28174.
PT New solid-phase competitive immunoassay using a fusion protein -
PT which binds to an antibody for the analyte and to an immobilised
PT antibody
PS Example 6; Column 23-24; 16pp; English.
CC R96579 is a portion of NP5, a fusion protein of beta-galactosidase and
CC an antigen from a human breast cancer cell line MCF-7, which binds
CC to monoclonal antibody Mc5 which binds to a breast epithelial mucin.
CC Mc5 only binds to 2 overlapping epitopes (see R96581 and R96582). The
CC fusion protein is used to demonstrate a new in vitro solid-phase,
CC competitive assay for detecting a peptide analyte in a sample. The
CC method uses two monoclonal antibodies which are respectively specific
CC for either a first or second peptide, in this case beta-galactosidase
CC and the breast cancer cell antigen. The method is partic. useful for

```

```

CC detecting the presence of a solid tumour or metastasis, as shown in
CC the above example, and also for the detection of growth hormone or
CC serum albumin.
SQ Sequence 93 AA;

Query Match 65.3%; Score 47; DB 18; Length 93;
Best Local Similarity 75.0%; Pred. No. 9.54e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 16 qgdrlrpgp 23
QY 13 QGHLRSGP 20

RESULT 4
ID R95022 standard; Protein: 3010 AA.
AC R95022;
DE Hepatitis GB virus (HGBV) contig C protein prod..
DE Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;
KW reagents; non-A; non-B; non-C; non-D; non-E; clone; GB contig C;
KW tamarin; infected plasma; lambda phage; cDNA library.
OS Hepatitis GB virus.
FH Key          Location/Qualifiers
FT misc_difference 1..3010
FT "note" "others correspond to degenerate or STOP
FT codons in T04247"
PN WO9521922-A2.
PD 17-AUG-1995.
PF 14-FEB-1995; U02118.
PR 14-FEB-1994; US-196030.
PR 13-MAY-1994; US-242654.
PR 29-JUL-1994; US-283314.
PR 23-NOV-1994; US-344190.
PR 23-NOV-1994; US-344185.
PR 27-JAN-1995; US-344557.
PA (ABBO ) ABBOTT LAB.
PI Buijk SL, Dawson GJ, Desai SM, Erker JC, Leary TP;
PI Muerhoff AS, Mushahwar IK, Pilot-Matias TJ, Schlauder GG;
PI Simons JN;
DR WPI: 95-293123/38.
DR N-PSDB: T04247.
PT Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful
PT for diagnosis and therapy of Hepatitis GB virus
PS Example 18; Pages 526-540; 661pp; English.
CC Double stranded hepatitis GB virus (HGBV) DNA obtd. from HGBV
CC infected tamarin plasma, using standard procedures, was used to
CC prepare a lambda phage HGBV cDNA library. Clones were rescued
CC from the lambda phage, searched against a sequence database and
CC found to be unique HGBV sequences. The clones were then used to
CC assemble the sequences GB contig A and B, which were amplified
CC using random primers. The prod. of which was amplified to give a
CC fragment of GB contig C, then using specified primers the
CC complete sequence of GB contig C was assembled to give T04247,
CC which encodes the proteins R95020-21 (the 3 possible coding strand
CC reading frames). Reagents which comprise the HGBV DNA, or its
CC protein prods. can be used for the diagnosis, therapy or in a
CC vaccine to prevent HGBV infection.
SQ Sequence 3010 AA;

Query Match 63.9%; Score 46; DB 16; Length 3010;
Best Local Similarity 63.8%; Pred. No. 1.25e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 2176 saghgrsqpk 2186
QY 12 AQGH-LRSGPR 21

RESULT 5
ID W18642 standard; Protein: 300 AA.
AC W18642;
DR 22-JUL-1997 (first entry)

```

DE Fragmented human amyloid A4 gene +2 frameshift mutant product.
 KW detection; antibody; probe; cancer; neoplasia; neurodegenerative;
 KW Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;
 KW Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
 KW PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;
 KW cardiovascular; rheumatoid arthritis.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 1..300
 FT /note= "X corresponds to a stop codon in the
 FT accompanying DNA file, T69789"
 FT peptide 179..188
 FT /note= "antigenic peptide used for antibody
 FT production"
 FT W09712992-A2.
 PN 10-APR-1997.
 PD 02-OCT-1996; IB1106.
 PR 02-OCT-1995; GB-020080.
 PR 11-JAN-1996; US-009832.
 PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
 PA (UYRO-) UNIV ROTTERDAM ERASMUS.
 PA (UYUT-) UNIV STATE UTRICHT.
 PI Burbach JPH, Grosveld FG, Van Leeuwen FW;
 DR WPI: 97-226235/20.
 DR N-PSDB; T69789.
 PT Use of mutant genes having frame:shift mutation(s) - for developing
 PT prods. for the diagnosis, prevention and treatment of associated
 PT diseases, e.g. cancer or neurodegenerative disease
 PS Claim 22; Fig 2; 123pp; English.
 CC W18642 and W18643 are +2 and +1 frameshift mutations, respectively, of a
 CC sequence comprising fragments of the coding sequence of the human amyloid
 CC A4 gene corresponding to nucleotides (nt) 147-586 followed by nt
 CC 1047-1526 of the wild-type amyloid A4 gene. These regions of the gene
 CC contain GAGAG motifs. Frameshift mutants of the tau, ubiquitin,
 CC apolipoprotein E, microtubule-associated protein 2 (MAP-2), neurofilament
 CC subunit L, M and H and amyloid A4 genes are claimed. All these genes
 CC share a common GAGAN motif (N= A, G, C or T), which is the site of common
 CC GA dinucleotide deletion(s) that cause neurodegenerative disorders.
 CC Antigenic peptides used for the production of antibodies, and small
 CC nucleic acid sequences derived from frameshift mutants are used in the
 CC diagnosis, prevention and treatment of cancer and neurodegenerative
 CC diseases, e.g. Parkinson's disease, Alzheimer's disease, Down's syndrome,
 CC frontal lobe dementia (Pick's disease), progressive supranuclear palsy
 CC (PSP), amyotrophic lateral sclerosis, Huntington's disease, multiple
 CC sclerosis, and other degenerative diseases such as cardiovascular
 CC disease and rheumatoid arthritis.
 SQ Sequence 300 AA;

Query Match 62.58; Score 45; DB 22; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.63e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 289 lrsqpr 294
 QY 16 LRSQPR 21
 IIIIIII

RESULT 6
 ID W40056 standard; Protein; 351 AA.
 AC W40056;
 DT 20-JUL-1998 (first entry)
 DE P300/CBP-associated transcriptional cofactor P/CAF N-terminus.
 KW P/CAF; human; P300; CBP; transcription; cofactor;
 KW histone acetyltransferase; HIV; infection; cancer; therapy;
 KW muscle differentiation.
 OS Homo sapiens.
 PN W09803652-A2.
 PD 29-JAN-1998.
 PR 23-JUL-1997; U12877.
 PR 23-JUL-1996; US-022273.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Howard BH, Nakatani Y;

DR WPI: 98-120777/11.
 PT New isolated p300/CBP-associated factor, P/CAF - used to develop
 PT products for modulating transcription, e.g. for treating HIV
 PT infection or cancers or for promoting muscle differentiation
 PS Claim 8; Page 74-75; 107pp; English.
 CC This polypeptide comprises the N-terminal region (amino acid
 CC residues 1-351) of a novel human p300/CBP associated cofactor,
 CC P/CAF (see W40052), that modulates transcription through binding
 CC to the cell transcription cofactors p300 (see W40055) and CBP
 CC (see W40060) and through acetylation of histones. This N-terminal
 CC region contains the binding site for p300/CBP. The invention
 CC provides methods of screening for compounds that inhibit or
 CC stimulate the transcription modulating and histone acetyltransferase
 CC activity of P/CAF and p300/CBP. Inhibitors can be used e.g. to
 CC inhibit HIV TAR-mediated transcription in the treatment of HIV
 CC infection. Stimulators can be used e.g. to activate tumour
 CC suppressor p53 in the treatment of cancer or to activate the muscle
 CC differentiation factor MyoD to promote muscle differentiation. The
 CC products can also be used to inhibit the cell cycle progression
 CC inducing effect of an oncoprotein which binds p300/CBP in a
 CC subject. The N-terminal region is specific for P/CAF and can be
 CC used to identify and define P/CAF.
 SQ Sequence 351 AA;

Query Match 61.1%; Score 44; DB 30; Length 351;
 Best Local Similarity 62.5%; Pred. No. 2.13e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 80 aqlrsapr 87
 QY 14 GHLRSGPR 21
 :IIII:II

RESULT 7
 ID W44155 standard; Protein; 529 AA.
 AC W44155;
 DT 14-MAY-1998 (first entry)
 DE Human neuronal nicotinic acetylcholine receptor alpha-2 subunit.
 KW Human; neuronal nicotinic acetylcholine receptor; alpha-2 subunit;
 KW brain tissue; screening; NACHR; antibody.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..55
 FT /label= signal
 FT Domain 264..289
 FT /label= TMD1
 FT /note= "transmembrane domain"
 FT 297..320
 FT Domain
 FT /label= TMD2
 FT /note= "transmembrane domain"
 FT 326..350
 FT Domain
 FT /label= TMD3
 FT /note= "transmembrane domain"
 FT 444..515
 FT Domain
 FT /label= TMD4
 FT /note= "transmembrane domain"
 FT 351..443
 FT Region
 FT /label= cytoplasmic_loop
 FT W09420617-A2.
 PN 15-SEP-1994.
 PD 08-MAR-1994; U02447.
 PR 08-MAR-1993; US-028031.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Elliott KJ, Ellis SB, Harpold MW;
 DR WPI: 94-303024/37.
 DR N-PSDB; V12199.
 PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -
 PT also transformed cells useful for screening cpds, which modulate
 PT activity of the receptor
 PS Disclosure; Page 68-70; 99pp; English.
 CC The present sequence represents a human neuronal nicotinic acetylcholine
 CC receptor (NACHR) subunit. The cells expressing the alpha and/or beta

CC NACHR subunits may be used in a method of screening compounds to
 CC identify any which modulate the activity of human neuronal NACHR.
 CC Subunit specific antibodies may be used to monitor the distribution
 CC and expression density of various subunits in normal vs diseased brain
 CC tissues. Testing of single receptor subunits or specific receptor
 CC subunit combinations with a variety of potential agonists or antagonists
 CC provides information with respect to the function and activity of the
 CC individual subunits and should lead to the identification and design of
 CC compounds that are capable of very specific interaction with one or
 CC more receptor subtypes. The resulting drugs should exhibit fewer
 CC unwanted side effects than drugs identified e.g. screening with cells
 CC that express a variety of subtypes.
 SQ Sequence 529 AA;

Query Match 61.1%; Score 44; DB 28; Length 529;
 Best Local Similarity 55.6%; Pred. No. 2.13e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 438 shghlhsga 446

QY 12 AQGHLRSGP 20
 :|||||:

RESULT 8

ID W09021 standard; Protein; 529 AA.
 AC W09021;
 DT 09-APR-1997 (first entry)
 DE Neuronal nicotinic acetylcholine receptor alpha-2 subunit.
 KW Neuronal nicotinic acetylcholine receptor; nACHR; neurotransmitter;
 KW ligand-gated receptor.
 OS Homo sapiens.
 PN W09641876-A1.
 PD 27-DEC-1996.
 PF 07-JUN-1996; U09775.
 PR 07-JUN-1995; US-484722.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Elliott KJ, Harpold MM;
 DR WPI: 97-065463/06.
 DR N-PSDB: T48235.
 PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -
 PT used in screening to determine the effect of drugs on the receptor
 PS Disclosure: Page 49-52; 108pp; English.
 CC The alpha-2 subunit (W09021) of the human neuronal nicotinic
 CC acetylcholine receptor (nACHR) can be expressed in transformed
 CC host cells carrying alpha-2 subunit DNA (see also T48235). Host
 CC cells, esp. mammalian cells or amphibian oocytes, expressing the
 CC recombinant alpha-2 subunit, opt. in combination with other
 CC recombinant alpha and/or beta subunits (see also W09018-20,
 CC W09022-27) can be used to examine the function of human AChR and
 CC to identify cpds. that modulate its activity.
 SQ Sequence 529 AA;

Query Match 61.1%; Score 44; DB 20; Length 529;
 Best Local Similarity 55.6%; Pred. No. 2.13e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 438 shghlhsga 446

QY 12 AQGHLRSGP 20
 :|||||:

RESULT 9

ID R73966 standard; Protein; 529 AA.
 AC R73966;
 DT 30-NOV-1995 (first entry)
 DE Alpha 2 subunit of a human neuronal nicotinic acetylcholine receptor.
 KW Human nACHR; neuronal nicotinic acetylcholine receptor;
 KW neurotransmitter.
 OS Homo sapiens.
 PN W09513299-A.
 PD 18-MAY-1995.
 PF 08-NOV-1994; U12859.
 PR 08-NOV-1993; US-149503.

PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Elliott KJ, Ellis SB, Harpold MM;
 DR WPI: 95-194036/25.
 DR N-PSDB: 090387.

PT New human neuronal nicotinic acetylcholine receptor alpha2 subunit DNA -
 PT used to develop prods. for detection, diagnosis and therapy and for
 PT modulating activity
 PS Disclosure: Page 43-46; 54pp; English.
 CC DNA encoding the human nACHR alpha2 subunit was isolated from
 CC human thalamus tissue cDNA library using corresp. rat cDNA. The
 CC insert of one clone obtd. was ligated with the insert of another
 CC clone to generate a full-length alpha 2 subunit cDNA. The DNA can be
 CC used to identify function nACHRs. Cells contg. the DNA can be used
 CC for screening to identify cpds. which modulate the activity of human
 CC nACHRs. The human nACHR alpha 2 subunit can be used to product
 CC antibodies which can be used in immunohistochemistry, diagnosis and
 CC therapy. The nucleic acids can be used for analysing disease states
 CC and creating animal models.
 SQ Sequence 529 AA;

Query Match 61.1%; Score 44; DB 14; Length 529;
 Best Local Similarity 55.6%; Pred. No. 2.13e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 438 shghlhsga 446

QY 12 AQGHLRSGP 20
 :|||||:

RESULT 10

ID W40059 standard; Protein; 813 AA.
 AC W40059;
 DT 20-JUL-1998 (first entry)
 DE Mouse p300/CBP-associated transcriptional cofactor P/CAF.
 KW P/CAF; mouse; p300; CBP; transcription; cofactor;
 KW histone acetyltransferase; HIV; infection; cancer; therapy;
 KW muscle differentiation.
 OS Mus musculus.
 PN W09803652-A2.
 PD 29-JAN-1998.
 PR 23-JUL-1997; U12877.
 PR 23-JUL-1996; US-022273.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Howard BH, Nakatani Y;
 DR WPI: 98-120777/11.
 DR N-PSDB: V10094.

PT New isolated p300/CBP-associated factor, P/CAF - used to develop
 PT products for modulating transcription, e.g. for treating HIV
 PT infection or cancers or for promoting muscle differentiation
 PS Disclosure: Page 86-88; 107pp; English.
 CC This polypeptide comprises a novel murine p300/CBP associated
 CC cofactor, designated P/CAF (see W40059). P/CAF modulates
 CC transcription through binding to the cell transcription cofactors
 CC p300 and CBP and through acetylation of histones. The invention
 CC provides methods of screening for compounds that inhibit or
 CC stimulate the transcription modulating and histone acetyltransferase
 CC activity of P/CAF (see also W40052) and p300/CBP. Such compounds
 CC can be used e.g. to treat HIV infection or to promote muscle
 CC differentiation. Antisense RNA or antisense DNA can be used to
 CC inhibit the activity of mouse P/CAF.
 SQ Sequence 813 AA;

Query Match 61.1%; Score 44; DB 30; Length 813;
 Best Local Similarity 62.5%; Pred. No. 2.13e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 62 aqlrsapr 69

QY 14 GHLRSGPR 21
 :|||||:

RESULT 11

ID W40052 standard; Protein; 832 AA.

AC W40052;
 DT 20-JUL-1998 (first entry)
 DE Human p300/CBP-associated transcriptional cofactor P/CAF.
 KW P/CAF; human; p300; CBP; transcription; cofactor;
 KW histone acetyltransferase; HIV; infection; cancer; therapy;
 KW muscle differentiation.
 OS Homo sapiens.
 FH Key
 FT Region 1..351
 FT /note= "contains binding site for p300/CBP
 FT (Claim 2)"
 FT Region 352..832
 FT /note= "contains histone acetyltransferase
 FT activity (Claim 8)"
 FT WO9803652-A2.
 PN 29-JAN-1998.
 PD 23-JUL-1997; U12877.
 PF 23-JUL-1996; US-022273.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Howard BH, Nakatani Y;
 PI WPI: 98-12077/11
 DR New isolated p300/CBP-associated factor, P/CAF - used to develop
 PT products for modulating transcription, e.g. for treating HIV
 PT infection or cancers or for promoting muscle differentiation
 PT Claim 1; Page 71-72; 107pp; English.
 PS This protein is a novel human p300/CBP associated cofactor,
 CC designated P/CAF, that modulates transcription through binding to
 CC the cell transcription cofactors p300 (see W40055) and CBP (see
 CC W40060) and through acetylation of histones. Its amino acid
 CC sequence was deduced from a clone (see V10090) obtained from a
 CC human foetal liver cDNA library. The N-terminal region (see
 CC W40056) of P/CAF, which contains the binding site for p300/CBP,
 CC and the C-terminal region (see W40054), which contains the histone
 CC acetyltransferase activity are also claimed. The invention provides
 CC methods of screening for compounds that inhibit or stimulate the
 CC transcription modulating and histone acetyltransferase activity of
 CC P/CAF and p300/CBP. Inhibitors can be used e.g. to inhibit HIV
 CC TAR-mediated transcription in the treatment of HIV infection.
 CC Stimulators can be used e.g. to activate tumour suppressor p53 in
 CC the treatment of cancer or to activate the muscle differentiation
 CC factor MyoD to promote muscle differentiation. The products can
 CC also be used to inhibit the cell cycle progression inducing effect
 CC of an oncoprotein which binds p300/CBP in a subject. The products
 CC can also be used for detection, screening assays and production of
 CC transgenic animals.
 SQ Sequence 832 AA;
 Query Match 61.1%; Score 44; DB 30; Length 832;
 Best Local Similarity 62.5%; Pred. No. 2.13e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 80 aqlrsapr 87
 QY 14 GHLRSGPR 21
 :|||:|
 RESULT 12
 ID W79294 standard; Protein; 1464 AA.
 AC W79294;
 DT 04-FEB-1999 (first entry)
 DE An antigen from dermonecrotic toxin of bronchiseptica.
 KW Antigen; dermonecrotic toxin; mixed vaccine; pig atrophic rhinitis;
 KW Pasteurella multocida toxin.
 OS Bordetella bronchiseptica.
 FH Key
 FT Location/Qualifiers
 FT Misc_difference 1
 FT /note= "encoded by GTC"
 FT J10251298-A.
 PN 22-SEP-1998.
 PD 31-JAN-1997; 053809.
 PF 01-FEB-1996; JP-050719.
 PR (HAND-) ZH HANDAI BISEIBUTSUBYO KENKYUKAI.
 PA WPI: 98-563160/48.
 DR

DR N-PSDB; V65398.
 PT Antigen comprising inactivated toxin of Bordetella pertussis -
 PT useful as a vaccine against, and a diagnostic agent for pig atrophic
 PT rhinitis
 PS Claim 1; Fig 1; 23pp; Japanese.
 CC The present sequence represents an antigen from the dermonecrotic toxin
 CC of Bordetella bronchiseptica. The antigen can be used as a mixed vaccine
 CC for pig atrophic rhinitis, by mixing Pasteurella multocida toxin with
 CC the antigen, or as a diagnostic agent for pig atrophic rhinitis in
 CC immunological assays.
 SQ Sequence 1464 AA;
 Query Match 61.1%; Score 44; DB 38; Length 1464;
 Best Local Similarity 50.0%; Pred. No. 2.13e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 582 pqgrfrtgr 591
 QY 12 AQGLRSGPR 21
 :|||:|
 RESULT 13
 ID P91333 standard; peptide; 36 AA
 AC P91333
 DT 19-MAR-1990 (first entry)
 DE Amino acid sequence of mod. Shiva- MP.
 KW Mod. shiva- MP; lytic peptide; antimicrobial peptide; disease-resistant
 KW trichophyte; shiva-1 MP.
 PN W08904371-A.
 PD 18-MAY-1989.
 PR 2-NOV-1987; US-115941.
 PA (LOUJ) Louisiana State Univ.
 PI Jaynes JM, Derrick KS;
 PI WPI: 89-165650/22.
 DR Transformed plants contg. heterologous gene - expressing antimicrobial
 PT Table 1; 56pp; English.
 PS Amino acid sequence of mod. shiva- MP as an exemplary lytic peptide for
 CC agent, or polypeptide high in essential amino acids
 CC use as an antimicrobial peptide contemplated for use in plant
 CC (trichophyte) transformants in the invention. Its amphiphilic homology
 CC with cecropin B is 100% and it is more lytically active than cecropin B.
 CC It is a homologue of shiva-1 SB.
 SQ Sequence 36 AA;
 Query Match 59.7%; Score 43; DB 1; Length 36;
 Best Local Similarity 75.0%; Pred. No. 2.77e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 17 qgilragp 24
 QY 13 QGHLRSGP 20
 :|||:|
 RESULT 14
 ID P91332 standard; peptide; 36 AA
 AC P91332
 DT 19-MAR-1990 (first entry)
 DE Amino acid sequence of Shiva-1 MP
 KW Shiva-1 MP; lytic peptide; antimicrobial peptide; disease-resistant
 KW trichophyte; cecropin B; mod. Shiva- MP.
 PN W08904371-A.
 PD 18-MAY-1989.
 PR 2-NOV-1987; US-115941.
 PA (LOUJ) Louisiana State Univ.
 PI Jaynes JM, Derrick KS;
 PI WPI: 89-165650/22.
 DR Transformed plants contg. heterologous gene - expressing antimicrobial
 PT Table 1; 56pp; English.
 PS Amino acid sequence of Shiva-1 MP as an exemplary lytic peptide for
 CC agent, or polypeptide high in essential amino acids
 CC use as an antimicrobial peptide contemplated for use in plant

CC (trichophyte) transformants in the invention. Its amphiphilic homology
 CC with cecropin B is 100% and it is more lytically active than cecropin B.
 CC Its lytic activity may be enhanced by removal or replacement of Gly-23
 CC and Pro-24. It is a homologue of mod. Shiva- SB.
 SQ Sequence 36 AA;

Query Match 59.7%; Score 43; DB 1; Length 36;
 Best Local Similarity 75.0%; Pred. No. 2.77e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 17 qgilragp 24
 |||||
 QY 13 QCHLRSGP 20

RESULT 15

ID R53354 standard; Protein; 63 AA.
 AC R53354;
 DT 24-SEP-1994 (first entry)
 DE sCEC-Shiva I fusion protein.
 KW Lytic peptide; resistance; fire blight; pomaceous fruit; scion;
 KW rootstock cultivar; transformation; transgenic variants; pears;
 KW apples; quinces; cecropin analogue.
 OS Hyalophora cecropia.
 PN W09407356-A.
 PD 14-APR-1994.
 PF 29-SEP-1993; U09299.
 PR 30-SEP-1992; US-954347.
 PR 18-MAR-1993; US-033772.
 PA (CORR) CORNELL RES FOUND INC.
 PI Aldwinckle HS, Norelli JL;
 DR WPI; 94-135092/16.
 DR N-PSDB; Q63459.
 PT Prodn of transgenic pomaceous fruit with fire blight resistance -
 PT by transforming fruit scion or rootstock cultivars with a gene
 PT encoding a lytic protein, e.g. lysozyme
 PS Disclosure: Page 13; 61pp; English.
 CC Resistance to fire blight conferred on pomaceous fruit scion or
 CC rootstock cultivars can be induced by transforming the plant with a
 CC gene encoding a lytic protein, e.g. lysozyme, cecropin, attacin or
 CC homologues. The transgenic variants show resistance to fire blight
 CC but retain their own characteristics. The method can be used to
 CC protect species of the Rosaceae family, e.g. pears, apples and
 CC quinces.
 CC See also R53346-58.
 SQ Sequence 63 AA;

Query Match 59.7%; Score 43; DB 10; Length 63;
 Best Local Similarity 75.0%; Pred. No. 2.77e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 44 qgilragp 51
 |||||
 QY 13 QCHLRSGP 20

Search completed: Thu Jul 8 18:07:40 1999
 Job time : 18 secs.

WQSEFH (TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:06:50 1999; Maspar time 4.90 Seconds
Tabular output not generated. 81.728 Million cell updates/sec

Title: >US-09-041-236-2
Description: (12-21) from US09041236.pep (2 of 45)
Perfect Score: 72
Sequence: 1 AQGHLRSGPR 10

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.450; Variance 26.906; scale 0.834

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	58	80.6	359	1	DEUT1B	4.67e-02
2	58	80.6	359	1	DEUT1C	4.67e-02
3	54	75.0	58	2	I53690	3.86e-01
4	54	75.0	383	2	S32975	3.86e-01
5	54	75.0	492	2	S27160	3.86e-01
6	52	72.2	57	2	I48627	1.07e+00
7	52	72.2	491	2	I48625	1.07e+00
8	52	72.2	491	2	A29818	1.07e+00
9	51	70.8	224	2	F70614	1.77e+00
10	51	70.8	498	1	JT0751	1.77e+00
11	50	69.4	534	2	J70842	2.91e+00
12	49	68.1	188	2	S09789	4.74e+00
13	48	66.7	416	2	A32947	7.69e+00
14	48	66.7	494	2	S11305	7.69e+00
15	47	65.3	342	2	S64042	1.24e+01
16	47	65.3	1083	2	S76111	1.24e+01
17	47	65.3	1191	2	S70963	1.24e+01
18	47	65.3	1774	2	B56101	1.24e+01
19	46	63.9	108	2	JQ2176	1.98e+01
20	46	63.9	118	2	S50866	1.98e+01
21	46	63.9	183	2	S33911	1.98e+01
22	46	63.9	622	2	A45746	1.98e+01
23	46	63.9	634	2	S35574	1.98e+01

24	46	63.9	743	2	T00634	hypothetical protein	1.98e+01
25	46	63.9	790	1	F0MVH2	gag-kit polyprotein p	1.98e+01
26	46	63.9	1212	2	T00332	hypothetical protein	3.14e+01
27	45	62.5	325	2	A64771	thiamin-phosphate kin	3.14e+01
28	45	62.5	345	2	E64115	thiamin-phosphate kin	3.14e+01
29	45	62.5	349	2	T01417	hypothetical protein	3.14e+01
30	45	62.5	491	2	A31047	testosterone 16alpha-	3.14e+01
31	45	62.5	491	2	I84735	testosterone 16a-hydr	3.14e+01
32	45	62.5	594	2	S39345	unc-18 protein homolo	3.14e+01
33	45	62.5	594	2	A33455	syntaxin-binding prot	3.14e+01
34	45	62.5	594	2	S39346	unc-18 protein homolo	3.14e+01
35	45	62.5	655	2	A42420	L-iduronidase (EC 3.2	3.14e+01
36	45	62.5	1191	2	A53491	bumetanide-sensitive	3.14e+01
37	45	62.5	1519	2	S41525	major ring-forming su	3.14e+01
38	45	62.5	1894	2	T02155	DNA-directed DNA poly	3.14e+01
39	44	61.1	216	2	C64993	hypothetical protein	4.94e+01
40	44	61.1	344	2	E70950	probable transposase	4.94e+01
41	44	61.1	367	2	C70858	hypothetical protein	4.94e+01
42	44	61.1	444	2	E65045	hypothetical protein	4.94e+01
43	44	61.1	711	2	S05381	VGFA protein precurs	4.94e+01
44	44	61.1	832	2	S71788	P/CAF protein - human	4.94e+01
45	44	61.1	1571	2	T00062	hypothetical protein	4.94e+01

ALIGNMENTS

RESULT 1
ENTRY DEUT1B #type complete
TITLE glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12),
ORGANISM glycosomal - Trypanosoma brucei
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 29-May-1998
ACCESSIONS S18806; S18807; S19071
REFERENCE
#authors Michels, P.A.M.; Poliszczak, A.; Osinga, K.A.; Misset, O.;
#journal van Beemmen, J.; Wierenga, R.K.; Borst, P.; Opperdoes, F.R.
#title EMBO J. (1986) 5:1049-1056
Two tandemly linked identical genes code for the glycosomal
glyceraldehyde-phosphate dehydrogenase in Trypanosoma
brucei.
#cross-references MUID:86247562
#accession S18806
#molecule_type DNA
#residues 1-359 #label MIC
#cross-references EMBL:X5955; NID:g11009; PID:g11010
#experimental_source strain 427
#genetics
#accession S18807
#molecule_type DNA
#residues 1-359 #label MTA
#cross-references EMBL:X5955; NID:g11009; PID:g11010
#experimental_source strain 427
#genetics
#accession S19071
#molecule_type protein
#residues 2-21, 'X', 23-48, 'X', 50-53 #label MIC2
GENETICS
#gene GAPDH1
GENETICS GAPDH2
#gene
CLASSIFICATION #superfamily glyceraldehyde-3-phosphate dehydrogenase
KEYWORDS glycolysis; glycosome; homotetramer; NAD; oxidoreductase
FEATURE
2-359 #product glyceraldehyde-3-phosphate dehydrogenase
#status experimental #label MAT
4-38 #region beta-alpha-beta NAD nucleotide-binding fold\
166,194 #active_site Cys, His #status predicted
SUMMARY
#length 359 #molecular-weight 39037 #checksum 3474
Query Match 80.6%; Score 58; DB 1; Length 359;
Best Local Similarity 70.0%; Pred. No. 4.67e-02; Indels 0; Gaps 0;
Matches 7; Conservative .3; Mismatches 0;

Db 120 AEGHLRGAR 129
1:||||:|
Qy 12 AOGHLRSGPR 21

RESULT 2
ENTRY DEUTIC #type complete
TITLE glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12),
glycosomal - Trypanosoma cruzi
ORGANISM #formal_name Trypanosoma cruzi
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
05-Sep-1997

ACCESSIONS S12565; S16508
REFERENCE S12565
#authors Kendall, G.; Wilderspin, A.F.; Ashall, F.; Miles, M.A.;
Kelly, J.M.
#journal EMBO J. (1990) 9:2751-2758
#title Trypanosoma cruzi glycosomal glyceraldehyde-3-phosphate
dehydrogenase does not conform to the 'hotspot' topogenic
signal model.
#cross-references MUID:90360985
#accession S12565
#molecule_type DNA
#residues 1-359 #label KEN1
#cross-references EMBL:X52898; NID:g10607; PID:g10608
#accession S16508
#molecule_type DNA
#residues 1-359 #label KEN2
#cross-references EMBL:X52898; NID:g10607; PID:g10608

GENETICS
#introns #status absent
#note tandem genes GAPDH I and II are identical within the
protein-coding region at the nucleotide level
CLASSIFICATION #superfamily glyceraldehyde-3-phosphate dehydrogenase
KEYWORDS gluconogenesis; glycolysis; glycosome; homotetramer; NAD;
oxidoreductase

FEATURE
4-38 #region beta-alpha-beta NAD nucleotide-binding fold\
166-194 #active_site Cys, His #status predicted
SUMMARY #length 359 #molecular-weight 39060 #checksum 3897

Query Match 80.6%; Score 58; DB 1; Length 359;
Best Local Similarity 70.0%; Pred. No. 4.67e-02;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 120 AEGHLRGAR 129
1:||||:|
Qy 12 AOGHLRSGPR 21

RESULT 3
ENTRY I53690 #type fragment
TITLE cytochrome P450 2B15 - rat (fragment)
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
05-Mar-1999

ACCESSIONS I53690
REFERENCE I53690
#authors Nakayama, K.; Suwa, Y.; Mizukami, Y.; Sogawa, K.;
Fujii-Kuriyama, Y.
#journal Gene (1993) 136:333-336
#title Cloning and sequencing of a novel rat cytochrome P450
2B-encoding gene.
#cross-references MUID:94124025
#accession I53690
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-58 #label RES
#cross-references GB:D17343; NID:g468474; PID:g468475

CLASSIFICATION #superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS heme; transmembrane protein

SUMMARY #length 58 #checksum 7474

Query Match 75.0%; Score 54; DB 2; Length 58;
Best Local Similarity 60.0%; Pred. No. 3.86e-01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 27 THGHLRPPGR 36
:||||:|
Qy 12 AOGHLRSGPR 21

RESULT 4
ENTRY S32975 #type complete
TITLE gene BCRF2 protein - human herpesvirus 4
ALTERNATE_NAMES gene BWRFL1 protein
ORGANISM #formal_name human herpesvirus 4, Epstein-Barr virus
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
13-Mar-1998

ACCESSIONS S32975; S32976; S32977; S32978; S32979; S32980; S32981;
S32982; S32983; S32984; S32985; S32986
REFERENCE S32973
#authors Farrell, P.J.
#submission submitted to the EMBL Data Library, March 1988
#accession S32975
#molecule_type DNA
#residues 1-383 #label FAR
#cross-references EMBL:V01555; NID:g59074; PID:e25077; PID:g1334836;
PID:e25078; PID:g1334837; PID:e25079; PID:g1334838;
PID:e25080; PID:g1334839; PID:e25081; PID:g1334840;
PID:e25066; PID:g1334841; PID:e25067; PID:g1334842;
PID:e25068; PID:g1334843; PID:e25069; PID:g1334844;
PID:e25070; PID:g1334845; PID:e25071; PID:g1334846;
PID:e25072; PID:g1334847

#note each of the twelve author-supplied translations in
EMBL:V01555 for this repeated gene is marked as
conflicting with the conceptual translation

GENETICS
#gene BCRF2_1; BWRFL_2; BWRFL_3; BWRFL_4; BWRFL_5; BWRFL_6;
BWRFL_7; BWRFL_8; BWRFL_9; BWRFL_10; BWRFL_11; BWRFL_12
#note twelve consecutive ORFs apparently encode the identical
polypeptide
SUMMARY #length 383 #molecular-weight 39866 #checksum 4975

Query Match 75.0%; Score 54; DB 2; Length 383;
Best Local Similarity 70.0%; Pred. No. 3.86e-01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 147 GQHLRVPPR 156
:||||:|
Qy 12 AOGHLRSGPR 21

RESULT 5
ENTRY S27160 #type complete
TITLE cytochrome P450 2B12 - rat
CONTAINS oxidoreductase (EC 1.-.-.-)
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
05-Mar-1999

ACCESSIONS S27160; S18907
REFERENCE S27160
#authors Friedberg, T.; Grassow, M.A.; Bartlomowicz-Oesch, B.;
Siegert, P.; Arand, M.; Adesnik, M.; Oesch, F.
#journal Biochem. J. (1992) 287:775-783
#title Sequence of a novel cytochrome CYP2B cDNA coding for a
protein which is expressed in a sebaceous gland, but not in
the liver.
#accession S27160
#molecule_type mRNA
#residues 1-492 #label FRI
#cross-references EMBL:X63545; NID:g56049; PID:g56050
#note the authors translated the codon CAT for residue 28 as
Arg and GAC for residue 83 as Gly

```
GENETICS
#gene
CLASSIFICATION CYP2B12
#superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS chromoprotein; electron transfer; endoplasmic reticulum;
heme; iron; monooxygenase; oxidoreductase; transmembrane
protein
FEATURE
437 #binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY #length 492 #molecular-weight 55796 #checksum 935
Query Match 75.0%; Score 54; DB 2; Length 492;
Best Local Similarity 60.0%; Pred. No. 3.86e-01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 27 THGLPPGPR 36
:||||:
QY 12 AQHLRSGPR 21

RESULT 6
ENTRY I49627 #type fragment
TITLE testosterone 16a-hydroxylase type c - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
05-Mar-1999
ACCESSIONS I49627
REFERENCE I49625
#authors Lakso, M.; Masaki, R.; Noshiro, M.; Negishi, M.
#journal Eur. J. Biochem. (1991) 195:477-486
#title Structures and characterization of sex-specific mouse
cytochrome P-450 genes as members within a large family.
Duplication boundary and evolution.
#cross-references MUID:91146586
#accession I49627
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-57 ##label RES
#cross-references GB:M60359; NID:g192889; PID:g192890
GENETICS
#gene
CLASSIFICATION 16aoh-c
#superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS heme; transmembrane protein
SUMMARY #length 57 #checksum 5864
Query Match 72.2%; Score 52; DB 2; Length 57;
Best Local Similarity 66.7%; Pred. No. 1.07e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 27 HGHLPPGPR 35
:||||:
QY 13 QGHLRSGPR 21

RESULT 7
ENTRY I49625 #type complete
TITLE testosterone 16a-hydroxylase type b - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
05-Mar-1999
ACCESSIONS I49625
REFERENCE I49625
#authors Lakso, M.; Masaki, R.; Noshiro, M.; Negishi, M.
#journal Eur. J. Biochem. (1991) 195:477-486
#title Structures and characterization of sex-specific mouse
cytochrome P-450 genes as members within a large family.
Duplication boundary and evolution.
#cross-references MUID:91146586
#accession I49625
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
```

```
##residues 1-491 ##label RES
#cross-references GB:M60358; NID:g192861; PID:g192863
GENETICS
#gene
CLASSIFICATION 16aoh-b
#superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS chromoprotein; heme; transmembrane protein
FEATURE
436 #binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY #length 491 #molecular-weight 55759 #checksum 9650
Query Match 72.2%; Score 52; DB 2; Length 491;
Best Local Similarity 66.7%; Pred. No. 1.07e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 27 HGHLPPGPR 35
:||||:
QY 13 QGHLRSGPR 21

RESULT 8
ENTRY A29818 #type complete
TITLE cytochrome P450 2B3, hepatic - rat
CONTAINS oxidoreductase (EC 1.-.-.-)
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change
05-Mar-1999
ACCESSIONS A29818; A25459
REFERENCE A29818
#authors Labbe, D.; Jean, A.; Anderson, A.
#journal DNA (1988) 7:253-260
#title A constitutive member of the rat cytochrome P450IIB
subfamily: full-length coding sequence of the P450IIB3
cDNA.
#cross-references MUID:88283344
#accession A29818
#molecule_type mRNA
#residues 1-491 ##label LAB
#cross-references GB:M20406; NID:g203683; PID:g203684
REFERENCE A25459
#authors Afolter, M.; Labbe, D.; Jean, A.; Raymond, M.; Noel, D.;
Labelle, Y.; Parent-Vaugeois, C.; Lambert, M.; Bojanowski,
R.; Anderson, A.
#journal DNA (1986) 5:209-218
#title cDNA clones for liver cytochrome P-450s from individual
aroclor-treated rats: constitutive expression of a new
P-450 gene related to phenobarbital-inducible forms.
#cross-references MUID:86246887
#accession A25459
#molecule_type mRNA
#residues 295-491 ##label AFF
GENETICS
#gene
CLASSIFICATION CYP2B3
#superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS chromoprotein; electron transfer; heme; iron; liver;
monooxygenase; oxidoreductase; transmembrane protein
FEATURE
436 #binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY #length 491 #molecular-weight 56384 #checksum 3521
Query Match 72.2%; Score 52; DB 2; Length 491;
Best Local Similarity 66.7%; Pred. No. 1.07e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 27 HGHLPPGPR 35
:||||:
QY 13 QGHLRSGPR 21
```

```

RESULT          9
ENTRY           F70614      #type complete
TITLE           probable malonyl coa-acyl carrier protein - Mycobacterium
                  tuberculosis (strain H37Rv)
ORGANISM        #formal_name Mycobacterium tuberculosis
DATE            17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                  17-Jul-1998
ACCESSIONS      F70614
REFERENCE       A70500
#authors        Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                  C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
                  III, C.E.; Tekala, E.; Badcock, K.; Basham, D.; Brown, D.;
                  Chillingworth, T.; Connor, R.; Davies, K.; Devlin, K.;
                  Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                  Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                  Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                  Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                  Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
                  Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal        Nature (1998) 393:537-544
#title          Deciphering the biology of Mycobacterium tuberculosis from
                  the complete genome sequence.
#cross-references MUID:98295987
#accession      F70614
#status         preliminary; nucleic acid sequence not shown;
                  translation not shown
#molecule_type DNA
#residues       1-224 #label COL
#cross-references GB:292772; GB:ALJ23456; NID:93261722; PID:e306578;
#experimental_source strain H37Rv
GENETICS
#gene           fadD37
SUMMARY         #length 224 #molecular_weight 23616 #checksum 5973
                  70.88; Score 51; DB 2; Length 224;
Query Match     Best Local Similarity 60.08; Pred. No. 1.77e+00;
Matches         6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db              41 AEARLRDGRP 50
QY              12 AOGHLRSGPR 21
                  |:::| | | |
RESULT          10
ENTRY           JT0751
TITLE           ferredoxin--NADP+ reductase (EC 1.18.1.2), long form
                  precursor - bovine
ALTERNATE_NAMES adrenodoxin reductase
ORGANISM         #formal_name Bos primigenius taurus #common_name cattle
DATE            14-Jul-1994 #sequence_revision 18-Oct-1996 #text_change
                  20-Mar-1998
ACCESSIONS      JT0751; JT0079; JS0390; S03558; PS0003; A29604; S52100
REFERENCE       JT0751
#authors        Takata, Y.; Sagara, Y.; Kono, A.; Sekimizu, K.; Horiuchi, T.
#journal        Biol. Pharm. Bull. (1993) 16:1200-1206
#title          Gene structure of bovine adrenodoxin reductase.
#cross-references MUID:94177140
#accession      JT0751
#molecule_type DNA
#residues       1-498 #label TAK
#experimental_source adrenal cortex
#note           the authors translated the codon GTC for residue 205 as
                  Gly
REFERENCE       JT0079
#authors        Sagara, Y.; Takata, Y.; Miyata, T.; Hara, T.; Horiuchi, T.
#journal        J. Biochem. (1987) 102:1333-1336
#title          Cloning and sequence analysis of adrenodoxin reductase cDNA
                  from bovine adrenal cortex.
#cross-references MUID:88198050
#accession      JT0079
#molecule_type mRNA
#residues       1-204,211-498 #label SAG

```

```

#cross-references GB:D00211; NID:9217433; PID:dl000595; PID:9217434
#note           the deduced sequence is partially confirmed by amino
                  acid sequencing of 15 isolated peptides
REFERENCE       JS0390
#authors        Sagara, Y.
#submission     submitted to DDBJ, September 1989
#revision       revision, insertion of residues 205-210
#accession      JS0390
#molecule_type mRNA
#residues       56-498 #label SA2
REFERENCE       S03558
#authors        Hanukoglu, I.; Gutfinger, T.
#journal        Eur. J. Biochem. (1989) 180:479-484
#title          cDNA sequence of adrenodoxin reductase. Identification of
                  NADP-binding sites in oxidoreductases.
#cross-references MUID:89170752
#accession      S03558
#molecule_type mRNA
#residues       135-204,211-498 #label HAN
#cross-references EMBL:X13736; NID:965; PID:9833776
#note           405-Ser was also found
REFERENCE       PS0003
#authors        Hamamoto, I.; Kurokohchi, K.; Tanaka, S.; Ichikawa, Y.
#journal        Biochim. Biophys. Acta (1988) 953:207-213
#title          Adrenoferradoxin-binding peptide of NADPH-adrenoferradoxin
                  reductase.
#cross-references MUID:88184054
#accession      PS0003
#molecule_type protein
#residues       33-41,'S',43-62;260-283,'TM',496-498 #label HAM
#note           a cyanogen bromide peptide binds to adrenoferradoxin
REFERENCE       A29604
#authors        Nonaka, Y.; Murakami, H.; Yabusaki, Y.; Kuramitsu, S.;
                  Kagamiyama, H.; Yamano, T.; Okamoto, M.
#journal        Biochim. Biophys. Res. Commun. (1987) 145:1239-1247
#title          Molecular cloning and sequence analysis of full-length cDNA
                  for mRNA of adrenodoxin oxidoreductase from bovine adrenal
                  cortex.
#cross-references MUID:87270696
#accession      A29604
#molecule_type mRNA
#residues       1-76,'R',78-80,'VWLALTPSRMLL',95-123,'RVYRLN',129-204,
                  211-273,'R',275-322,'RL',325-328,'ARRSAMOSPE',340-346,
                  'HPSAHWGCGP',359-498 #label NON
#cross-references GB:M17029; NID:6162628; PID:g162629
#experimental_source adrenal cortex
REFERENCE       S52100
#authors        Warburton, R.J.; Seybert, D.W.
#journal        Biochim. Biophys. Acta (1995) 1246:39-46
#title          Structural and functional characterization of bovine
                  adrenodoxin reductase by limited proteolysis.
#cross-references MUID:95110846
#accession      S52100
#status         preliminary
#molecule_type protein
#residues       'X',34-41,'X',43-48,'X',50-51;304-306,'X',308-309,'X',
                  311-326 #label WAR
COMMENT         Ferredoxin--NADP+ reductase is localized in the matrix of adrenal
                  cortex mitochondria and functions as a component of the
                  mitochondrial steroid-hydroxylating enzyme system, which includes
                  ferredoxin--NADP+ reductase, adrenodoxin and two forms of
                  cytochrome P-450.
GENETICS
#introns        27/1: 59/3; 91/3; 132/3; 170/3; 204/3; 246/3; 275/1: 341/3;
                  359/1: 456/1
CLASSIFICATION #superfamily human ferredoxin--NADP+ reductase
KEYWORDS        alternative splicing; flavoprotein; mitochondrion; monomer;
                  NADP; oxidoreductase
FEATURE         #domain transit peptide (mitochondrion) #status
                  predicted #label SIG\
                  1-32
                  #product ferredoxin--NADP+ reductase, long form #status
                  predicted #label MAT\
                  33-498

```

33-204,211-498 #product ferredoxin--NADP+ reductase, short form #status experimental #label MA2\
40-70 #region beta-alpha-beta FAD nucleotide-binding fold\
180-190 #region NAD binding #status Predicted\
281 #binding_site substrate (Lys) #status experimental
SUMMARY #length 498 #molecular-weight 55008 #checksum 7201

Query Match 70.8%; Score 51; DB 1; Length 498;
Best Local Similarity 87.5%; Pred. No. 1.77e+00;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 435 GHLPSPGR 442

QY 14 GHLSRSGP 21

RESULT 11

ENTRY TITLE C70842 #type complete
probable pmmb protein - Mycobacterium tuberculosis (strain H37Rv)

ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998

ACCESSIONS C70842

REFERENCE A70500

#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.

#cross-references MUID:98295987
#accession C70842
#status preliminary; nucleic acid sequence not shown;
translation not shown

#molecule_type DNA
#residues 1-534 #label COL
#cross-references GB:AL021841; GB:AL123456; NID:g3261517; PID:e1251139;
PID:g2894217

#experimental_source strain H37Rv

GENETICS

#gene pmmb

SUMMARY #length 534 #molecular-weight 56195 #checksum 8368

Query Match 69.4%; Score 50; DB 2; Length 534;
Best Local Similarity 75.0%; Pred. No. 2.91e+00;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 43 RGHLSGSP 50

QY 13 QGHLSRSGP 20

RESULT 12

ENTRY TITLE S09789 #type complete
hypotheoretical protein UL26 - human cytomegalovirus (strain AD169)

ORGANISM #formal_name human cytomegalovirus, human herpesvirus 5
#note host Homo sapiens (man)

DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Sep-1997

ACCESSIONS S09789

REFERENCE S09749

#authors Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.; Hutchinson III, C.A.; Kouzarides, T.; Martignetti, J.A.; Preddie, E.; Satchwell, S.C.;

Tomlinson, P.; Weston, K.M.; Barrell, B.G.
Curr. Top. Microbiol. Immunol. (1990) 154:125-169
#title Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169.

#cross-references MUID:90269039

#accession S09789

#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-188 #label CHE
#cross-references EMBL:X17403; NID:g59591; PID:g59631
#note this sequence was submitted to the EMBL Data Library, December 1989

SUMMARY #length 188 #molecular-weight 21155 #checksum 3949

Query Match 68.1%; Score 49; DB 2; Length 188;
Best Local Similarity 70.0%; Pred. No. 4.74e+00;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 128 AQGLRHGMR 137

QY 12 AQGHLRSGP 21

RESULT 13

ENTRY A32947 #type fragment

TITLE filaggrin precursor - human (fragment)

ORGANISM #formal_name Homo sapiens #common_name man

DATE 20-Dec-1989 #sequence_revision 04-Sep-1992 #text_change 24-Sep-1998

ACCESSIONS A32947

REFERENCE A32947

#authors McKinley-Grant, L.J.; Idler, W.W.; Bernstein, I.A.; Parry, D.A.D.; Cannizzaro, L.; Croce, C.M.; Huebner, K.; Lessin, S.R.; Steinert, P.M.

#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:4848-4852

#title Characterization of a cDNA clone encoding human filaggrin and localization of the gene to chromosome region 1q21.

#cross-references MUID:89296901

#accession A32947

#status preliminary

#molecule_type mRNA

#residues 1-416 #label MCK

#cross-references GB:M24355; NID:g182604; PID:g182605

#note the authors translated the codon CAC for residue 188 as Gln, and AAT for residue 307 as Gln

GENETICS

#gene GDB:FLG

#cross-references GDB:119912; OMIM:135940

#map_position 1q21-1q21

CLASSIFICATION #superfamily unassigned calmodulin-related proteins;

calmodulin repeat homology

epidermis; polymorphism; tandem repeat

SUMMARY #length 416 #checksum 9609

Query Match 66.7%; Score 48; DB 2; Length 416;
Best Local Similarity 60.0%; Pred. No. 7.89e+00;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 207 AQGLRDGSR 216

QY 12 AQGHLRSGP 21

RESULT 14

ENTRY S11305 #type complete

TITLE cytochrome P450 2B11 - dog

ALTERNATE_NAMES cytochrome P450IIB

CONTAINS oxidoreductase (EC 1.-.-.-)

ORGANISM #formal_name Canis lupus familiaris #common_name dog

DATE 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 05-Mar-1999

ACCESSIONS S11305

REFERENCE S11305

#authors Graves, P.E.; Elhag, G.A.; Ciaccio, P.J.; Bourque, D.P.;
#journal Halpert, J.R. Biochem. Biophys. (1990) 281:106-115
#title cDNA and deduced amino acid sequences of a dog hepatic
cytochrome P450IIB responsible for the metabolism of 2,2',
4,4',5,5'-hexachlorobiphenyl.
#cross-references MUID:90343348
#accession S11305
#molecule_type mRNA
##residues 1-494 ##label GRA
##cross-references EMBL:M92447; NID:g164039; PID:g164040
GENETICS
#gene CYP2B11
CLASSIFICATION #superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS chromoprotein; electron transfer; endoplasmic reticulum;
heme; iron; monooxygenase; oxidoreductase; phosphoprotein;
transmembrane protein
FEATURE
436 #binding_site heme iron (Cys) (axial ligand) #status
Predicted
SUMMARY #length 494 #molecular-weight 56266 #checksum 6061
Query Match 66.7%; Score 48; DB 2; Length 494;
Best Local Similarity 75.0%; Pred. No. 7.59e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 28 GHLPPGPR 35
||| :|||
Qy 14 GHLRSGPR 21
RESULT 15
ENTRY S64042 #type complete
TITLE porphobilinogen synthase (EC 4.2.1.24) - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES aminolevulinate dehydratase; protein G3610; protein YGL040c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change
06-Feb-1998
ACCESSIONS S64042; S64044; A28465
REFERENCE S64003
#authors Hebling, U.; Hofmann, B.; Delius, H.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64042
#molecule_type DNA
##residues 1-342 ##label HEB
##cross-references EMBL:272562; NID:g1322523; PID:e243278; PID:g1322524;
##experimental_source strain S288C
REFERENCE S64044
#authors Feuerhann, M.; Potier, S.; Souciet, J.L.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64044
#molecule_type DNA
##residues 41-342 ##label FEU
##cross-references EMBL:272562; MIPS:YGL040c
##experimental_source strain S288C
REFERENCE A28465
#authors Myers, A.M.; Crivellone, M.D.; Koerner, T.J.; Tzagoloff, A.
#journal J. Biol. Chem. (1987) 262:16822-16829
#title Characterization of the yeast HEM2 gene and transcriptional
regulation of COX5 and COR1 by heme.
#cross-references MUID:88059078
#accession A28465
#molecule_type DNA
##residues 1-290,'D',292-342 ##label MYE
##cross-references EMBL:J03493; NID:g171663; PID:g171664
GENETICS
#gene SGD:HEM2
#map_position 7L
CLASSIFICATION #superfamily porphobilinogen synthase

KEYWORDS carbon-oxygen lyase; hydro-lyase; porphyrin biosynthesis
SUMMARY #length 342 #molecular-weight 37740 #checksum 6043
Query Match 65.3%; Score 47; DB 2; Length 342;
Best Local Similarity 66.7%; Pred. No. 1.24e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 316 QGFLRAGAR 324
||| :|||
Qy 13 QGHLRSGPR 21

Search completed: Thu Jul 8 18:07:00 1999
Job time : 10 secs.

WIRE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:05:31 1999; MasPar time 3.66 Seconds
Tabular output not generated. 77.243 Million cell updates/sec

Title: >US-09-041-236-2
Description: (12-21) from US09041236.pep (2 of 45)
Perfect Score: 72
Sequence: 1 AQHLSRGPR 10

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 23.106; Variance 24.254; scale 0.953

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	58	80.6	1	G3PG_TRYBB	9.60e-03
2	58	80.6	1	G3PG_TRYR	9.60e-03
3	54	75.0	1	CPB3_RAT	1.01e-01
4	52	72.2	1	CPB3_BOVIN	3.14e-01
5	51	70.8	1	ADRO_BOVIN	5.49e-01
6	49	68.1	1	UL26_HCMVA	1.84e+00
7	48	66.7	1	G3P_LYOSH	2.79e+00
8	48	66.7	1	FILA_HUMAN	2.79e+00
9	48	66.7	1	CPB3_CANFA	2.79e+00
10	47	65.3	1	HEM2_YEAST	4.73e+00
11	47	65.3	1	CPB3_MOUSE	4.73e+00
12	47	65.3	1	NIFJ_RHOU	4.73e+00
13	47	65.3	1	CAIH_MOUSE	4.73e+00
14	46	63.9	1	YHUS_YEAST	7.94e+00
15	46	63.9	1	GAG_FSVHZ	7.94e+00
16	46	63.9	1	ALU8_HUMAN	7.94e+00
17	46	63.9	1	PPCC_HUMAN	7.94e+00
18	45	62.5	1	THIL_SALTY	1.32e+01
19	45	62.5	1	THIL_ECOLI	1.32e+01
20	45	62.5	1	THIL_HAEIN	1.32e+01
21	45	62.5	1	CPB9_MOUSE	1.32e+01
22	45	62.5	1	STX1_BOVIN	1.32e+01
23	45	62.5	1	STX1_MOUSE	1.32e+01

24	45	62.5	594	1	STX1_RAT	1.32e+01
25	45	62.5	655	1	IDUA_CANFA	1.32e+01
26	45	62.5	1108	1	CYGE_MOUSE	1.32e+01
27	45	62.5	1191	1	NKCL_SQUAC	1.32e+01
28	44	61.1	124	1	V124_ASFLS	2.18e+01
29	44	61.1	235	1	GRPE_MYCTU	2.18e+01
30	44	61.1	444	1	YGAF_ECOLI	2.18e+01
31	44	61.1	444	1	FRE5_MOUSE	2.18e+01
32	44	61.1	445	1	MRS4_ECOLI	2.18e+01
33	44	61.1	529	1	ACH2_HUMAN	2.18e+01
34	44	61.1	673	1	UN18_CAEEL	2.18e+01
35	44	61.1	756	1	YD34_SCHPO	2.18e+01
36	44	61.1	1021	1	BUB1_YEAST	2.18e+01
37	43	59.7	70	1	YAON_RHISN	3.55e+01
38	43	59.7	169	1	YS4L_PNECA	3.55e+01
39	43	59.7	283	1	MAUR_PARDE	3.55e+01
40	43	59.7	442	1	GAG_VILV2	3.55e+01
41	43	59.7	442	1	GAG_VILV	3.55e+01
42	43	59.7	593	1	ALU6_HUMAN	3.55e+01
43	43	59.7	605	1	APM2_YEAST	3.55e+01
44	43	59.7	1111	1	MYSB_DICDI	3.55e+01
45	43	59.7	1596	1	GLI3_HUMAN	3.55e+01

ALIGNMENTS

RESULT 1
ID G3PG_TRYBB STANDARD; PRT; 358 AA.
AC P22512;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, GLYCOSOMAL (EC 1.2.1.12)
DE (GAPDH).
OS TRYPANOSOMA BRUCEI BRUCEI.
RC EUKARYOTA; EUKLENOZA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 1-52.
RX MEDLINE; 86247562.
RA MICHELIS P.A.M., POLISZCZAK A., OSINGA K.A., MISSET O.,
RA VAN BUEMAN J., WIEGENGA R.K., BORST P., OPPERDOES F.R.;
RT "Two tandemly linked identical genes code for the glycosomal
RT glyceraldhyde-phosphate dehydrogenase in Trypanosoma brucei."
RL EMO J. 5:1049-1056(1986).
CC -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
CC + NAD(+) -> 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -!- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: GLYCOSOMAL.
CC -!- THERE ARE TWO IDENTICAL GENES THAT CODES FOR GLYCOSOMAL GAPDH.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
CC or send an email to license@isb-sib.ch).

EMBL; X59955; G11010; -
DR EMBL; X59955; G11011; -
EMBL; M26816; G162089; -
DR PIR; S18806; DEUT1B.
DR PIR; S18807; S18807.
PDB; 1GGA; 31-JAN-94.
DR PROSITE; PS00071; GAPDH; 1.
DR PROSITE; PS00342; MICROBODIES_CTER; 1.
DR PFAM; PF00044; gpdh; 1.
KW GLYCOLYSIS; OXIDOREDUCTASE; NAD; GLYCOSOME; MULTIGENE FAMILY;
3D-STRUCTURE.
FT INIT_MET 0 165
FT BINDING 165 165 GLYCERALDEHYDE 3-PHOSPHATE.

FT ACT_SITE 193 193 ACTIVATES THIOL GROUP DURING CATALYSIS.
FT SITE 356 358 MICROBODY TARGETING SIGNAL (POTENTIAL).

DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, GLYCOSOMAL (EC 1.2.1.12)
DE (GAPDH).
OS TRYPANOSOMA CRUZI.
OC EUKARYOTA; EUKLENOZOA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=X10/6;
RX MEDLINE; 90360985.
RA KENDALL G., WILDERSPIN A.F., ASHALL F., MILES M.A., KELLY J.M.;
RT "Trypanosoma cruzi glycosomal glyceraldehyde-3-phosphate
RT dehydrogenase does not conform to the 'hotspot' topogenic signal
RT model";
RL EMBO J. 9:2751-2758(1990).
CC -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
CC + NAD(+) -> 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -!- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: GLYCOSOMAL.
CC -!- THERE ARE TWO IDENTICAL GENES THAT CODES FOR GLYCOSOMAL GAPDH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52898; G10609; -
CC EMBL; X52898; G10608; -
CC PIR; S12565; DEUTIC.
CC PIR; S16508; S16508.
CC PROSITE; PS00071; GAPDH; 1.
CC PFAM; PF00044; gpdh; 1.
CC HSP; P22512; IGGA.
KW GLYCOLYSIS; OXIDOREDUCTASE; NAD; GLYCOSOME; MULTIGENE FAMILY.
FT BINDING 166 166 GLYCERALDEHYDE 3-PHOSPHATE.
FT ACT_SITE 194 194 ACTIVATES THIOL GROUP DURING CATALYSIS.
FT SITE 357 359 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 359 AA; 39060 MW; 4A32B52A CRC32;

Query Match 80.6%; Score 58; DB 1; Length 359;
Best Local Similarity 70.0%; Pred. No. 9.60e-03;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 120 AEGHLRGGAR 129
QY 12 AQHLRSRGP 21
|:||||:|

RESULT 3
ID CPBC RAT STANDARD; PRT; 492 AA.
AC P33272;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 2B12 (EC 1.14.14.1) (CYP1B12).
GN CYP2B12 OR CYP2B-12.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE; 93075030.
RA FRIEDBERG T., GRASSOW M.A., BARTLOMOWICZ-OESCH B., SIEGERT P.,
RA ARAND M., ADESNIK M., OESCH F.;
RT "Sequence of a novel cytochrome CYP2B cDNA coding for a protein which
RT is expressed in a sebaceous gland, but not in the liver";
RL BIOCHEM. J. 287:775-783(1992)
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. THIS ISOZYME SEEMS

Query Match 80.6%; Score 58; DB 1; Length 358;
Best Local Similarity 70.0%; Pred. No. 9.60e-03;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 119 AEGHLRGGAR 128
QY 12 AQHLRSRGP 21
|:||||:|

RESULT 2
ID G3PG-TRYCR STANDARD; PRT; 359 AA.
AC P22513;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)

CC RESPONSIBLE FOR METABOLISM OF 2,2',4,4',5,5'-HEXACHLOROBIPHENYL.
CC -|- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2O).
CC -|- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -|- TISSUE SPECIFICITY: PREPUTIAL GLAND, BUT NOT IN LIVER.
CC -|- TISSUE SPECIFICITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63545; G56050; -.
DR PIR: S18907; S8907.
DR PIR: S27160; S27160.
DR PFAM: PF00067; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;
KW MICROSOME; ENDOPLASMIC RETICULUM.
FT BINDING 437 437 HEME (BY SIMILARITY).
SQ SEQUENCE 492 AA; 55796 MW; 4A06129A CRC32;
Query Match 75.0%; Score 54; DB 1; Length 492;
Best Local Similarity 60.0%; Pred. No. 1.01e-01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 27 THGLPPGPR 36
QY 12 AQHLRSGPR 21
:||||:||||
RESULT 4
ID CYP2B3-RAT STANDARD; PRT; 491 AA.
AC P13107;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 13-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 2B3 (EC 1.14.14.1) (CYP2B3).
GN CYP2B3 OR CYP2B-3.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88283344.
RA LABBE D., JEAN A., ANDERSON A.;
RT "A constitutive member of the rat cytochrome P450IIB subfamily: full-
RT length coding sequence of the P450IIB3 cDNA.";
RL DNA 7:253-260(1988).
CC -|- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATY
CC ACIDS, AND XENOBIOTICS.
CC -|- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2O).
CC -|- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -|- TISSUE SPECIFICITY: LIVER; NOT FOUND IN THE LUNG, KIDNEY, AND
CC PROSTATE.
CC -|- INDUCTION: CONSTITUTIVELY EXPRESSED.
CC -|- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: M20406; G203684; -.
DR PIR: A29818; A29818.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR PFAM: PF00067; P450; 1.
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;
KW MICROSOME; ENDOPLASMIC RETICULUM.
FT BINDING 436 436 HEME.
SQ SEQUENCE 491 AA; 56384 MW; D97FA331 CRC32;
Query Match 72.2%; Score 52; DB 1; Length 491;
Best Local Similarity 66.7%; Pred. No. 3.14e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 27 HGLPPGPR 35
QY 13 QGHLRSGPR 21
:||||:||||
RESULT 5
ID ADRO_BOVIN STANDARD; PRT; 492 AA.
AC P08165;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NADPH:ADRENODOXIN OXIDOREDUCTASE PRECURSOR (EC 1.18.1.2) (ADRENODOXIN
DE REDUCTASE) (FERREDOXIN-NADP(+) REDUCTASE).
GN FDXR OR ADXR.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94177140.
RA TAKATA Y., SAGARA Y., KONO A., SEKIMIZU K., HORIUCHI T.;
RT "Gene structure of bovine adrenodoxin reductase.";
RL BIOL. PHARM. BULL. 16:1200-1206(1993).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 86198050.
RA SAGARA Y., TAKATA Y., MIYATA T., HARA T., HORIUCHI T.;
RT "Cloning and sequence analysis of adrenodoxin reductase cDNA from
RT bovine adrenal cortex.";
RL J. BIOCHEM. 102:1333-1336(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87270696.
RA NONAKA Y., MURAKAMI H., YABUSAKI Y., KURAMITSU S., KAGAMIYAMA H.,
RA YAMANO T., OKAMOTO M.;
RT "Molecular cloning and sequence analysis of full-length cDNA for mRNA
RT of adrenodoxin oxidoreductase from bovine adrenal cortex.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 145:1239-1247(1987).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=ADRENAL CORTEX;
RX MEDLINE; 89170752.
RA HANUKOGLU I., GUTINGER T.;
RT "cDNA sequence of adrenodoxin reductase. Identification of
RT NADP-binding sites in oxidoreductases.";
RL EUR. J. BIOCHEM. 180:479-484(1989).
RN [5]
RP ALTERNATIVE SPLICING.
RA SAGARA Y.;
RL SUBMITTED (SEP-1989) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -|- CATALYTIC ACTIVITY: REDUCED ADRENODOXIN + NADP(+) = OXIDIZED
CC ADRENODOXIN + NADPH.
CC -|- COFACTOR: FAD FLAVOPROTEIN.
CC -|- PATHWAY: CHOLESTEROL SIDE-CHAIN-CLEAVAGE SYSTEM.
CC -|- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -|- ALTERNATIVE PRODUCTS: AN ALTERNATIVELY SPLICED MRNA CONTAINING 6
CC EXTRA RESIDUE REPRESENTS 10-20% OF ALL ADRENODOXIN REDUCTASE MRNA.
CC THIS LONGER FORM SEEMS TO BE INACTIVE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: M17029; G162629; -;
DR EMBL: D00211; D1000595; -;
DR DR EMBL: X13736; G833776; -;
DR PIR: A29604; A29604.
DR PIR: JS0390; JS0390.
DR PIR: S03558; S03558.
DR PIR: JT0751; JT0751.
DR KW OXIDOREDUCTASE; FLAVOPROTEIN; NADP; FAD; MITOCHONDRION;
KW TRANSIT PEPTIDE; ALTERNATIVE SPLICING.
FT TRANSIT 1 32 MITOCHONDRION.
FT CHAIN 33 492 ADRENODOXIN REDUCTASE.
FT VARSPLIC 204 204 E -> EVLLLCQ (IN ALTERNATIVE SPLICED
FT INACTIVE VARIANT).
FT CONFLICT 77 77 G -> R (IN REF. 3).
FT CONFLICT 81 94 FGVPADHPVKVNI -> VWLALTTPRSRMLL (IN REF.
FT 3).
FT CONFLICT 124 128 QDAYH -> RVYRLT (IN REF. 3).
FT CONFLICT 268 268 K -> R (IN REF. 3).
FT CONFLICT 317 318 PS -> RL (IN REF. 3).
FT CONFLICT 323 333 RAAGIRLAVTR -> ARRSAWOSPE (IN REF. 3).
FT CONFLICT 341 352 TRAVPTGDVEDL -> HPGSAHWGCGP (IN REF. 3).
SQ SEQUENCE 492 AA; 54338 MW; 656F1A2B CRC32;

Query Match 70.8%; Score 51; DB 1; Length 492;
Best Local Similarity 87.5%; Pred. No. 5.49e-01;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 429 GHLPSGPR 436

QY 14 GHLRSRGP 21

RESULT 6
ID UL26_HCMVA STANDARD; PRT; 188 AA.
AC P16762;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN UL26.
GN UL26.
OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90269039.
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTIGNETTI J.A.,
RA PREDDIE E., SATCHELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC -1- SIMILARITY: BELONGS TO US22 FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: X17403; G59631; -;
DR PIR: S09789; S09789.
KW HYPOTHETICAL PROTEIN.
FT CARBOHYD 136 136 POTENTIAL.

SQ SEQUENCE 188 AA; 21155 MW; 45076C6C CRC32;
Query Match 68.1%; Score 49; DB 1; Length 188;
Best Local Similarity 70.0%; Pred. No. 1.64e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 128 AQRRLRHGNNR 137
QY 12 AQRHLRSRGP 21

RESULT 7
ID G3P_LYOSH STANDARD; PRT; 337 AA.
AC Q92243;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH).
GN GPD.
OS LYOPHYLLUM SHIMEJI.
OC EUKARYOTA; FUNGI; BASIDIOMYCOTA; HYMENOMYCETES; AGARICALES;
OC TRICHOLOMATACEAE; LYOPHYLLUM.
RN [1]
RP SEQUENCE FROM N.A.
RA SAITO T., TANAKA N., TOYOMASU T.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: D8426; G1655498; -;
DR PROSITE: PS00071; GAPDH; 1.
DR PFAM: PF00044; gpdh; 1.
DR HSP: P00354; 3GPD.
KW GLYCOLYSIS; OXIDOREDUCTASE; NAD.
FT BINDING 150 150 GLYCERALDEHYDE 3-PHOSPHATE.
FT ACT_SITE 177 177 ACTIVATES THIOL GROUP DURING CATALYSIS.
SQ SEQUENCE 337 AA; 36724 MW; 99D4E555 CRC32;

Query Match 66.7%; Score 48; DB 1; Length 337;
Best Local Similarity 50.0%; Pred. No. 2.79e+00;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 106 AEGHLKGGAK 115
QY 12 AQRHLRSRGP 21

RESULT 8
ID FILA_HUMAN STANDARD; PRT; 416 AA.
AC P20930;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE FILAGGRIN PRECURSOR (FRAGMENT).
GN FLG.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89296901.
RA MCKINLEY-GRANT L.J., IDLER W.W., BERNSTEIN I.A., PARRY D.A.D.,

RA CANNIZZARO L., CROCE C.M., HUEBNER K., LESSIN S.R., STEINERT P.M.;
RT "Characterization of a cDNA clone encoding human filaggrin and
RL localization of the gene to chromosome region 1q21.";
CC PROC. NATL. ACAD. SCI. U.S.A. 86:4848-4852(1989).
CC -!- FUNCTION: FILAGGRIN AGGREGATES KERATIN INTERMEDIATE FILAMENTS
CC AND PROMOTES DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE
CC FILAMENTS DURING TERMINAL DIFFERENTIATION OF MAMMALIAN
CC EPIDERMIS.
CC -!- P.TM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
CC OF 324 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
CC PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M24355; G182605; -
DR PIR: A32947; A32947.
DR MIM: 135940; -
KW PHOSPHORYLATION: POLYPROTEIN; DEVELOPMENTAL PROTEIN.
FT NON_TER 1 1
SQ SEQUENCE 416 AA; 44105 MW; 72D2B913 CRC32;

Query Match 66.7%; Score 48; DB 1; Length 416;
Best Local Similarity 60.0%; Pred. No. 2.79e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 207 AQQLRDGSR 216
|| :|| :||
QY 12 AQGLRSGPR 21

RESULT 9
ID CPBB_CANFA STANDARD; PRT; 494 AA.
AC P24450;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 2B11 (EC 1.14.14.1) (CYPI1B1) (P450 PBD-2).
GN CYP2B1.
OS CANIS FAMILIARIS (DOG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BEAGLE; TISSUE=LIVER;
RX MEDLINE: 90343348.
RA GRAVES P.E., ELRHAG G.A., CIACCIO P.J., BOURQUE D.P., HALPERT J.R.;
RT "cDNA and deduced amino acid sequences of a dog hepatic cytochrome
RT P450IIB responsible for the metabolism of 2,2',4,4',5,5'-
RT hexachlorobiphenyl.";
RL ARCH. BIOCHEM. BIOPHYS. 281:106-115(1990).
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. THIS ISOZYME SEEMS
CC RESPONSIBLE FOR METABOLISM OF 2,2',4,4',5,5'-HEXACHLOROBIPHENYL.
CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -!- INDUCTION: BY PHENOBARBITAL.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M92447; G164040; -
DR PIR: S11305; S11305.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR PFAM: PF00067; p450; 1.
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;
KW MICROSOME; ENDOPLASMIC RETICULUM; PHOSPHORYLATION.
FT MOD_RES 128 128 PHOSPHORYLATION (BY CAPK) (BY
FT SIMILARITY).
SQ BINDING 436 436 HEME (BY SIMILARITY).
SQ SEQUENCE 494 AA; 56266 MW; 4BA10BBE CRC32;

Query Match 66.7%; Score 48; DB 1; Length 494;
Best Local Similarity 75.0%; Pred. No. 2.79e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 28 GHLPPGPR 35
||| :|||
QY 14 GHLRSGPR 21

RESULT 10
ID HEM2_YEAST STANDARD; PRT; 342 AA.
AC P05373;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DELTA-AMINOLEVULINIC ACID DEHYDRATASE (EC 4.2.1.24) (PORPHOBILINOGEN
DE SYNTHASE) (ALADH).
GN HEM2 OR YGL040C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88059078.
RA MYERS A.M., CRIVELLONE M.D., KOERNER T.J., TZAGOLOFF A.;
RT "Characterization of the yeast HEM2 gene and transcriptional
RT regulation of COX5 and COR1 by heme.";
RL J. BIOL. CHEM. 262:16822-16829(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA HEBLING U., HOFMANN B., DELIUS H.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 1-41 FROM N.A.
RC STRAIN=S288C;
RX MEDLINE: 973777993.
RA FEUERMAN M., DE MONTIGNY J., POTIER S., SOUCIET J.-L.;
RT "The characterization of two new clusters of duplicated genes
RT suggests a 'Lego' organization of the yeast Saccharomyces cerevisiae
RT chromosomes.";
RL YEAST 13:861-869(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE: 98069651.
RA ERSKINE P.T., SENIOR N., AWAN S., LAMBERT R., LEWIS G., TICKLE I.J.,
RA SARWAR M., SPENCER P., THOMAS P., WARREN M.J., SHOOLINGIN-JORDAN P.M.,
RA WOOD S.P., COOPER J.B.;
RT "X-ray structure of 5-aminolaevulinic acid dehydratase, a hybrid
RT aldolase.";
RL NAT. STRUCT. BIOL. 4:1025-1031(1997).
CC -!- CATALYTIC ACTIVITY: 2 5-AMINOLEVULINATE = PORPHOBILINOGEN +
CC 2 H(2)O.
CC -!- COFACTOR: ZINC.
CC -!- PATHWAY: SECOND STEP IN PORPHYRIN AND HEME BIOSYNTHESIS.
CC -!- SUBUNIT: HOMOOCTAMER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: J03493; G171664; -;
DR EMBL: 272562; E243278; -;
DR PIR: A28465; A28465.
DR PDB: 1AW5; 18-NOV-98.
DR SGD: L0000761; HEM2.
DR PROSITE: PS00169; D_ALA_DEHYDRATASE; 1.
DR PFAM: PF00490; ALAD; 1.
KW PORPHYRIN BIOSYNTHESIS; HEME BIOSYNTHESIS; LYASE; ZINC; 3D-STRUCTURE.
FT DOMAIN 130 148
FT ACT_SITE 263 263
FT CONFLICT 291 291
SQ SEQUENCE 342 AA; 37740 MW; 708D318A CRC32;

Query Match 65.3%; Score 47; DB 1; Length 342;
Best Local Similarity 66.7%; Pred. No. 4.73e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 316 QGFLRAGAR 324
QY 13 QGHLRSGPR 21

RESULT 11
ID CPBJ_MOUSE STANDARD; PRT; 492 AA.
AC 055071.
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 2B19 (EC 1.14.14.1) (CYP2B19).
GN CYP2B9.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
[1]

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=SKIN;
RX MEDLINE; 99017986.
RA KENEY D.S.:
RT "The novel skin-specific cytochrome P450 Cyp2b19 maps to proximal
chromosome 7 in the mouse, near a cluster of Cyp2 family genes.";
RL GENOMICS 53:417-419(1998).

CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
ACIDS, AND XENOBIOTICS.

CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
OXIDIZED FLAVOPROTEIN + H(2)O.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -!- TISSUE SPECIFICITY: EXPRESSED ONLY IN DIFFERENTIATED KERATINOCYTES
IN SKIN.

CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: AF047529; G2920650; -;
DR MGD: MGI:107303; CYP2B19.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;
FT MICROsome; ENDOPLASMIC RETICULUM; PHOSPHORYLATION.
FT MOD_RES 129 129 PHOSPHORYLATION (BY CAPK) (BY
SIMILARITY).

FT BINDING 437 437 HEME (BY SIMILARITY).
SQ SEQUENCE 492 AA; 55996 MW; E5C3DF9B CRC32;

Query Match 65.3%; Score 47; DB 1; Length 492;
Best Local Similarity 50.0%; Pred. No. 4.73e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 THGHFFPGPR 36
QY 12 AQGLRSGPR 21

RESULT 12
ID NIFJ_RHORU STANDARD; PRT; 1191 AA.
AC Q53046;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PYRUVATE-FLAVODOXIN OXIDOREDUCTASE (EC 1.-.-.-).
GN NIFJ.
OS RHODOSPIRILLUM RUBRUM.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOSPIRILLACEAE;
OC RHODOSPIRILLUM.
OC [1]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 1170;
RX MEDLINE; 96347128.
RA LINDBLAD A., JANSSON J., BROSTEDT E., JOHANSSON M., HELLMAN U.,
RA NORDLUND S.:
RT "Identification and sequence of a nifJ-like gene in Rhodospirillum
rubrum: partial characterization of a mutant unaffected in nitrogen
fixation.";
RT fixation.";
RL MOL. MICROBIOL. 20:559-568(1996).

CC -!- FUNCTION: OXIDOREDUCTASE REQUIRED FOR THE TRANSFER OF ELECTRONS
FROM PYRUVATE TO FLAVODOXIN, WHICH REDUCES NITROGENASE.
CC -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: X77515; G453436; -;
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; PARTIAL.
DR PFAM: PF00037; fer4; 1.
KW OXIDOREDUCTASE; NITROGEN FIXATION; ELECTRON TRANSPORT; IRON-SULFUR;
4FE-4S.

FT METAL 696 696 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 699 699 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 702 702 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 706 706 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 753 753 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 756 756 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 759 759 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 763 763 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
SQ SEQUENCE 1191 AA; 127829 MW; 319B6090 CRC32;

Query Match 65.3%; Score 47; DB 1; Length 1191;
Best Local Similarity 75.0%; Pred. No. 4.73e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 474 SHLRFQPR 481
QY 14 GHLRSGPR 21

RESULT 13
ID CAIH_MOUSE STANDARD; PRT; 1315 AA.

AC P39061;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE COLLAGEN ALPHA 1(XVIII) CHAIN PRECURSOR.
 GN COL18A1.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE=LIVER;
 RX MEDLINE: 94245707.
 RA REHN M.V., HINTIKKA E., PIHLAJANIEMI T.;
 RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
 RT partial structure of the corresponding gene, and comparison of the
 RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
 RT chain.";
 RL J. BIOL. CHEM. 269:13929-13935(1994).
 RN [2]
 RP SEQUENCE OF 1-928 FROM N.A.
 RX MEDLINE: 94240112.
 RA REHN M.V., PIHLAJANIEMI T.;
 RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
 RT collagenous sequence, a distinct tissue distribution, and homology
 RT with type XV collagen.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 91:4234-4238(1994).
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L16998; G553894; -
 DR EMBL: U03714; G487734; -
 DR EMBL: U03715; G1167905; -
 DR EMBL: U34606; G1167905; JOINED.
 DR EMBL: U34608; G1167905; JOINED.
 DR EMBL: U34609; G1167905; JOINED.
 DR EMBL: U34610; G1167905; JOINED.
 DR EMBL: U34611; G1167905; JOINED.
 DR EMBL: U34612; G1167905; JOINED.
 DR EMBL: U34613; G1167905; JOINED.
 DR EMBL: U03716; G1167905; JOINED.
 DR EMBL: U03718; G1167905; JOINED.
 DR EMBL: U11636; G618428; -
 DR EMBL: MGI:88451; COL18A1.
 DR HSP: P20081; 1YAT.
 DR GSD: MGI:88451; COL18A1.
 KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
 KW CELL ADHESION; COLLAGEN; GLYCOPROTEIN; SIGNAL.
 FT SIGNAL
 FT CHAIN 1 25 POTENTIAL.
 FT DOMAIN 26 1315 COLLAGEN ALPHA 1(XVIII) CHAIN.
 FT DOMAIN 26 326 NONHELICAL REGION 1 (NC1).
 FT DOMAIN 327 333 TRIPLE-HELICAL REGION 2 (COL1).
 FT DOMAIN 354 363 NONHELICAL REGION 2 (NC2).
 FT DOMAIN 364 437 TRIPLE-HELICAL REGION 2 (COL2).
 FT DOMAIN 438 461 TRIPLE-HELICAL REGION 3 (NC3).
 FT DOMAIN 462 583 TRIPLE-HELICAL REGION 3 (COL3).
 FT DOMAIN 584 606 NONHELICAL REGION 4 (NC4).
 FT DOMAIN 607 689 TRIPLE-HELICAL REGION 4 (COL4).
 FT DOMAIN 690 703 NONHELICAL REGION 5 (NC5).
 FT DOMAIN 704 745 TRIPLE-HELICAL REGION 5 (COL5).
 FT DOMAIN 746 738 NONHELICAL REGION 6 (NC6).
 FT DOMAIN 759 831 TRIPLE-HELICAL REGION 6 (COL6).
 FT DOMAIN 832 841 NONHELICAL REGION 7 (NC7).
 FT DOMAIN 842 874 TRIPLE-HELICAL REGION 7 (COL7).
 FT DOMAIN 875 886 NONHELICAL REGION 8 (NC8).

FT DOMAIN 887 910 TRIPLE-HELICAL REGION 8 (COL8).
 FT DOMAIN 911 917 NONHELICAL REGION 9 (NC9).
 FT DOMAIN 918 969 TRIPLE-HELICAL REGION 9 (COL9).
 FT DOMAIN 970 982 NONHELICAL REGION 10 (NC10).
 FT DOMAIN 983 1000 TRIPLE-HELICAL REGION 10 (COL10).
 FT DOMAIN 1001 1315 NONHELICAL REGION 11 (NC11).
 FT CARBOHYD 126 126 POTENTIAL.
 FT CARBOHYD 488 488 POTENTIAL.
 FT SITE 892 894 CELL ATTACHMENT SITE (POTENTIAL).
 SQ SEQUENCE 1315 AA; 134263 MW; 134263 CRC32;
 Query Match 65.3%; Score 47; DB 1; Length 1315;
 Best Local Similarity 50.0%; Pred. No. 4.73e+00;
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Db 1221 SOGOLQPGAR 1230
 QY 12 AQGLRSGPR 21
 :||:|:|:|:
 RESULT 14
 ID YH08_YEAST STANDARD; PRT; 183 AA.
 AC P32899;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE PUTATIVE MITOCHONDRIAL 40S RIBOSOMAL PROTEIN YHRI48W.
 GN YHRI48W.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
 OC SACCHAROMYCETACEAE; SACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93365023.
 RA SCHWANK S., HARRER R., SCHUELLER H.-J., SCHWEIZER E.;
 RT "Molecular cloning and analysis of the nuclear gene MRP-L6 coding for
 RT a putative mitochondrial ribosomal protein from Saccharomyces
 RT cerevisiae.";
 RL CURR. GENET. 24:136-140(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RX MEDLINE: 94378003.
 RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
 RA DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,
 RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
 RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
 RA NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,
 RA VIGNATI D., WILCOX L., WORLDMAN P., WATERSTON R., WILSON R.,
 RA VAUDIN M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT VIII.";
 RL SCIENCE 265:2077-2082(1994).
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X69480; G4014; -
 DR EMBL: U10397; G500654; -
 DR PIR: S28572; S28572.
 DR PIR: S33911; S33911.
 DR PIR: S46758; S46758.
 DR PROSITE: PS00632; RIBOSOMAL_S4; FALSE_NEG.
 DR PFAM: PF00163; S4; 1.
 KW HYPOTHETICAL PROTEIN; RIBOSOMAL PROTEIN; MITOCHONDRION.
 SQ SEQUENCE 183 AA; 21885 MW; CD420920 CRC32;

Query Match 63.9%; Score 46; DB 1; Length 183;
 Best Local Similarity 75.0%; Pred. No. 7.94e+00;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 133 QGHVRVGP 140
 QY 13 QGHLRSGP 20

```

RESULT 15
ID GAG_FSVHZ STANDARD; PRT; 414 AA.
AC P04322.
DT 20-MAR-1987 (REL. 04, CREATED)
DT 20-MAR-1987 (REL. 04, LAST SEQUENCE UPDATE)
DT 01-JAN-1988 (REL. 06, LAST ANNOTATION UPDATE)
DE GAG POLYPROTEIN [CONTAINS: CORE PROTEINS P15, P12, P30] (FRAGMENT).
GN GAG.
OS FELINE SARCOMA VIRUS (STRAIN HARDY-ZUCKERMAN 4).
OC VIRUSES; RETROVIRIDAE; RETROVIRIDAE; MAMMALIAN TYPE C RETROVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86175044.
RA BESMER P., MURPHY J.E., GEORGE P.C., OIU F., BERGOLD P.J.,
RA LEDERMAN L., SNYDER H.W. JR., BRODEUR D., ZUCKERMAN E.E., HARDY W.D.;
RT "A new acute transforming feline retrovirus and relationship of its
RT oncogene v-kit with the protein kinase gene family.";
RL NATURE 320:415-421(1986).
CC -!- THIS PROTEIN IS SYNTHESIZED AS A GAG-KIT-POL POLYPROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03711; G61536; ALT_TERM.
DR PIR; A03936; FOMVHZ.
DR PFAM; PF01140; gag_MA; 1.
DR PFAM; PF01141; gag_P12; 1.
KW CORE PROTEIN; POLYPROTEIN; MYRISTYLATION.
PROPEP 1 74 LEADER PEPTIDE.
FT CHAIN 75 201 CORE PROTEIN P15.
FT CHAIN 202 271 CORE PROTEIN P12.
FT CHAIN 272 >414 CORE PROTEIN P30.
FT LIPID 76 76 MYRISTATE.
FT NON_TER 414 414
SQ SEQUENCE 414 AA; 45732 MW; AE2A54C4 CRC32;

Query Match 63.9%; Score 46; DB 1; Length 414;
Best Local Similarity 60.0%; Pred. No. 7.94e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 237 GHGPLPSGPR 246
QY 12 AQGHLRSGPR 21
    
```

Search completed: Thu Jul 8 18:05:43 1999
 Job time : 12 secs.

W P E R L (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:09:39 1999; Maspar time 6.33 seconds
Tabular output not generated. 43.704 Million cell updates/sec

Title: >US-09-041-236-2
Description: (25-37) from US09041236.pap (3 of 45)
Perfect Score: 108
Sequence: 1 VWKGHVQDRVDF 13

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-gene3seq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 19.443; Variance 59.365; scale 0.328

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	65	60.2	403	32	W59641	Amino acid sequence o 7.44e+00
2	63	58.3	402	32	W59642	Amino acid sequence o 1.22e+01
3	61	56.5	673	26	W22502	Phaffia derived carot 2.00e+01
4	61	56.5	673	26	W22500	Phaffia derived carot 2.00e+01
5	61	56.5	673	26	W22497	Phaffia derived carot 2.00e+01
6	61	56.5	903	28	W35006	Polyangium brachyspor 2.00e+01
7	60	55.6	1206	30	W47030	Bovine N-proteinase. 2.56e+01
8	59	54.6	1066	39	W88255	Morchella costata alp 3.26e+01
9	59	54.6	1066	38	R70638	Morchella alpha-1,4-g 3.26e+01
10	59	54.6	1066	13	R72711	Alpha-1,4-glucan lyas 3.26e+01
11	57	52.8	419	16	R79930	Porcine acylglucosami 5.28e+01
12	56	51.9	348	27	W36499	Hereditary haemochrom 6.70e+01
13	55	50.9	291	30	W52352	Bacillus subtilis pro 8.50e+01
14	55	50.9	319	36	W49875	Pyrococcus VCI-7Eg1 g 8.50e+01
15	55	50.9	319	28	W35007	Pyrococcus furiosus e 8.50e+01
16	55	50.9	319	35	W29729	Pyrococcus furiosus e 8.50e+01

17	55	50.9	1120	39	W81642	Mouse elf protein.	8.50e+01
18	55	50.9	2154	39	W81639	Mouse elf-1 protein.	8.50e+01
19	53	49.1	369	22	W20779	H. pylori flagella-as	1.36e+02
20	53	49.1	495	34	W49070	Streptococcus pneumon	1.36e+02
21	53	49.1	607	6	R32201	Full length grapevine	1.36e+02
22	52	48.1	402	16	R79928	Porcine acylglucosami	1.72e+02
23	52	48.1	402	16	R79931	Porcine acylglucosami	1.72e+02
24	52	48.1	417	16	R79929	Porcine acylglucosami	1.72e+02
25	52	48.1	494	3	P70443	Sequence encoded by r	1.72e+02
26	52	48.1	506	39	W83126	PrtIIK48 lysine speci	1.72e+02
27	52	48.1	1732	17	R96029	P. gingivalis porphyr	1.72e+02
28	52	48.1	1732	24	W24787	prtk antigenic protei	1.72e+02
29	52	48.1	1732	36	W69487	Haemagglutinin protei	1.72e+02
30	51	47.2	13	4	P51245	Sequence that corresp	2.16e+02
31	51	47.2	16	3	R15369	Ig idiotypic determin	2.16e+02
32	51	47.2	16	4	P51246	Sequence that corresp	2.16e+02
33	51	47.2	121	10	R54797	SPA-reactive IgM heav	2.16e+02
34	51	47.2	220	27	W23681	French bean polypheno	2.16e+02
35	51	47.2	365	33	W55967	Human stress-activat	2.16e+02
36	51	47.2	365	23	W23783	Human mitogen-activat	2.16e+02
37	51	47.2	365	25	W26578	Human MAP kinase homo	2.16e+02
38	50	46.3	181	13	R65933	Human CBF-beta.	2.72e+02
39	50	46.3	185	25	W26488	KSHV DNA polymerase.	2.72e+02
40	50	46.3	196	27	W23664	Strawberry polyphenol	2.72e+02
41	50	46.3	205	4	R20934	Envelope proteins fro	2.72e+02
42	50	46.3	390	39	W82675	Soil bacteria peptide	2.72e+02
43	50	46.3	735	26	W22729	Z. mays starch syntha	2.72e+02
44	50	46.3	816	13	R66931	AMML chromosome inv(1	2.72e+02
45	50	46.3	885	13	R66930	AMML chromosome inv(1	2.72e+02

ALIGNMENTS

RESULT 1
ID W59641 standard; Protein; 403 AA.
AC W59641;
DT 24-SEP-1998 (first entry)
DE Amino acid sequence of mouse Stac protein.
KW Mouse; Stac protein; mstac; cysteine rich domain; SH3 domain;
KW neuron information transfer.
FH Key Location/Qualifiers
FT Domain 108..160
FT Domain /note= "cysteine rich domain"
FT Domain 293..340
FT Domain /note= "SH3 domain"
PN J10175998-A.
PD 30-JUN-1998.
PF 13-DEC-1996; 332606.
PR 13-DEC-1996; JP-332606.
PR (SHIO) SHIONOGI & CO LTD.
DR WPI; 98-422394/36.
DR N-PSDB; V41436.
PT Protein comprising cysteine rich and SH3 domain(s) - has important
PT role in facilitating neuron information transfer
PS Disclosure; Pages 7-11; 17pp; Japanese.
CC This is the amino acid sequence of the mouse Stac (mStac) protein,
CC comprising a cysteine rich, and SH3 domains. The protein has an
CC important role in neuron information transfer.
SQ Sequence 403 AA;

Query Match 60.2%; Score 55; DB 32; Length 403;
Best Local Similarity 66.7%; Pred. No. 7.44e+00;
Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Db 324 wkiki-qdrvgf 334
Qy 26 WKGHVQDRVDF 37

RESULT 2
ID W59642 standard; Protein; 402 AA.
AC W59642;

DT 24-SEP-1998 (first entry)
 DE Amino acid sequence of human Stac protein.
 KW Human; Stac protein; hstac; cysteine rich domain; SH3 domain;
 KW neuron information transfer.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Domain 109..161 /note="cysteine rich domain"
 FT Domain 294..341 /note="SH3 domain"
 FT
 PN J10175998-A.
 PD 30-JUN-1998.
 PF 13-DEC-1996; 332606.
 PR 13-DEC-1996; JP-332606.
 PA (SHIO) SHIONOGI & CO LTD.
 DR WPI: 98-422394/36.
 DR N-PSDB; V41437.
 PT Protein comprising cysteine rich and SH3 domain(s) - has important
 PT role in facilitating neuron information transfer
 PS Claim 1; Pages 11-14; 17pp; Japanese.
 CC This is the amino acid sequence of the human stac (hstac) protein,
 CC comprising a cysteine rich, and SH3 domains. The protein has an
 CC important role in neuron information transfer.
 SQ Sequence 402 AA;

Query Match 58.3%; Score 63; DB 32; Length 402;
 Best Local Similarity 58.3%; Pred. No. 1.22e+01;
 Matches 7; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Db 323 wkqkl-qdrigf 333
 ||| : ||| : |
 QY 26 WKGHVQDRYDF 37

RESULT 3
 ID W22502 standard; Protein; 673 AA.
 AC W22502;
 DT 10-MAR-1998 (first entry)
 DE Phaffia derived carotenoid biosynthesis pathway enzyme PRGcrtB GB.
 KW Glyceraldehyde-3-phosphate dehydrogenase; GAPDH; carotenoid;
 KW synthesis; promoter; recombinant DNA; astaxanthin; ribosomal protein;
 KW food colouring.
 OS Phaffia rhodozyma.
 PN WO9723633-A1.
 PD 03-JUL-1997.
 PF 23-DEC-1996; E05887.
 PR 11-APR-1996; EP-200943.
 PR 22-DEC-1995; EP-203620.
 PA (KONN) GIST-BROCADES BV.
 PI Verdoes JC, Wery J;
 PI WPI: 97-351068/32.
 DR N-PSDB; T72946.
 PT Phaffia derived GAPDH and carotenoid synthesis genes and promoter
 PT fragment - used in the recombinant production of therapeutically
 PT useful proteins e.g. carotenoids for use in food colouring
 PS Claim 11; Page 72-74; 118pp; English.
 CC The present sequence represents a Phaffia derived carotenoid
 CC biosynthesis pathway enzyme. The nucleic acid encoding this protein
 CC can be used in the novel recombinant DNA of the present invention. The
 CC recombinant DNA comprises a transcription promoter operably linked to a
 CC downstream sequence to be expressed, where the transcription promoter
 CC comprises a region found upstream of the open reading frame (ORF) of a
 CC highly expressed Phaffia gene (preferably GAPDH, a ribosomal protein
 CC gene, an enzyme involved in the biosynthesis pathway). The recombinant
 CC DNA can be used to transform hosts, preferably Phaffia. These
 CC transformed hosts are then used in the recombinant production of GAPDH
 CC or an enzyme involved in carotenoid synthesis, preferably astaxanthin.
 CC They may also be used to produce a pharmaceutical product. Purified
 CC carotenoids can be used as colourants in food and/or feed, and also in
 CC cosmetics. 673 AA;

Query Match 56.5%; Score 61; DB 26; Length 673;
 Best Local Similarity 70.0%; Pred. No. 2.00e+01;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 642 vvkqdvger 651
 ||||| ||| : |
 QY 25 VMKGHVQDR 34
 ||||| ||| : |
 RESULT 4
 ID W22500 standard; Protein; 673 AA.
 AC W22500;
 DT 10-MAR-1998 (first entry)
 DE Phaffia derived carotenoid biosynthesis pathway enzyme PRcrtY.
 KW Glyceraldehyde-3-phosphate dehydrogenase; GAPDH; carotenoid;
 KW synthesis; promoter; recombinant DNA; astaxanthin; ribosomal protein;
 KW food colouring.
 OS Phaffia rhodozyma.
 PN WO9723633-A1.
 PD 03-JUL-1997.
 PF 23-DEC-1996; E05887.
 PR 11-APR-1996; EP-200943.
 PR 22-DEC-1995; EP-203620.
 PA (KONN) GIST-BROCADES BV.
 PI Verdoes JC, Wery J;
 PI WPI: 97-351068/32.
 DR N-PSDB; T72944.
 PT Phaffia derived GAPDH and carotenoid synthesis genes and promoter
 PT fragment - used in the recombinant production of therapeutically
 PT useful proteins e.g. carotenoids for use in food colouring
 PS Claim 11; Page 63-65; 118pp; English.
 CC The present sequence represents a Phaffia derived carotenoid
 CC biosynthesis pathway enzyme. The nucleic acid encoding this protein
 CC can be used in the novel recombinant DNA of the present invention. The
 CC recombinant DNA comprises a transcription promoter operably linked to a
 CC downstream sequence to be expressed, where the transcription promoter
 CC comprises a region found upstream of the open reading frame (ORF) of a
 CC highly expressed Phaffia gene (preferably GAPDH, a ribosomal protein
 CC gene, an enzyme involved in the biosynthesis pathway). The recombinant
 CC DNA can be used to transform hosts, preferably Phaffia. These
 CC transformed hosts are then used in the recombinant production of GAPDH
 CC or an enzyme involved in carotenoid synthesis, preferably astaxanthin.
 CC They may also be used to produce a pharmaceutical product. Purified
 CC carotenoids can be used as colourants in food and/or feed, and also in
 CC cosmetics. 673 AA;

Query Match 56.5%; Score 61; DB 26; Length 673;
 Best Local Similarity 70.0%; Pred. No. 2.00e+01;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 642 vvkqdvger 651
 ||||| ||| : |
 QY 25 VMKGHVQDR 34

RESULT 5
 ID W22497 standard; Protein; 673 AA.
 AC W22497;
 DT 10-MAR-1998 (first entry)
 DE Phaffia derived carotenoid biosynthesis pathway enzyme PRcrtB.
 KW Glyceraldehyde-3-phosphate dehydrogenase; GAPDH; carotenoid;
 KW synthesis; promoter; recombinant DNA; astaxanthin; ribosomal protein;
 KW food colouring.
 OS Phaffia rhodozyma.
 PN WO9723633-A1.
 PD 03-JUL-1997.
 PF 23-DEC-1996; E05887.
 PR 11-APR-1996; EP-200943.
 PR 22-DEC-1995; EP-203620.
 PA (KONN) GIST-BROCADES BV.
 PI Verdoes JC, Wery J;
 PI WPI: 97-351068/32.

KW anhydrofructose; beverage; wine; foodstuff.
OS Morchella costata.
PN WO9850532-A2.
PD 12-NOV-1998.
PF 06-MAY-1998; IBO708.
PR 06-MAY-1997; GB-009161.
PA (DANI-) DANISCO AS.
PI Buchter-Larsen A, Marcussen I;
DR WPI: 99-070094/06.
DR N-PSDB: V84194.
PT Recombinant production of anti-oxidant compounds - by the production
PT of an anhydrofructose from a glucan, used for improving plants for
PT use in foodstuffs.
PS Claim 7; Page 36-38; 53pp; English.
CC This is the amino acid sequence of an alpha-1,4-glucan lyase of
CC Morchella costata that is used in claimed processes of the
CC invention. A claimed process of preparing a medium that contains
CC an antioxidant and at least one other component involves preparing
CC antioxidant from a glucan, in situ, by use of recombinant DNA
CC techniques. Also claimed are: (1) use of anhydrofructose as an
CC antioxidant for a medium, where the anhydrofructose is prepared in
CC situ in the medium; (2) use of anhydrofructose for imparting or
CC improving stress tolerance in a plant, where the anhydrofructose is
CC prepared in situ in the plant; (3) use of glycan lyase (see
CC W88253-57 and W88278) for imparting or improving: (i) stress
CC tolerance in plant, or (ii) transformation of grape, where the
CC glucan lyase is prepared in situ; (4) use of a nucleotide sequence
CC (NS) coding for a glucan lyase (see V84192-97) as a means of
CC imparting or improving stress tolerance in a plant, where the NS is
CC expressed in situ in the plant, and (5) use of a NS coding for a
CC glucan lyase for imparting or improving the transformation of grape,
CC where the NS is expressed in situ in the grape. The antioxidant
CC containing medium can be used as a foodstuff or in the preparation
CC of foodstuffs such as beverages, in particularly alcoholic
CC beverages such as wine (claimed).
SQ Sequence 1066 AA;

Query Match 54.6%; Score 59; DB 39; Length 1066;
Best Local Similarity 41.7%; Pred. No. 3.26e+01;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 891 wtgggganrikf 902
| | : | : | : |
QY 26 WKGHVGQDRVDF 37

RESULT 9
ID R70638 standard; Protein; 1066 AA.
AC R70638;
DT 10-JAN-1996 (first entry)
DE M.costata alpha-1,4-glucan lyase.
KW Alpha-1,4-glucan lyase; 1,5-D-anhydrofructose; G.lemaleiformis;
KW antioxidant; food additive; sweetener.
OS Morchella costata.
PN WO9510616-A2.
PD 20-APR-1995.
PF 15-OCT-1994; E03397.
PR 15-OCT-1993; GB-021302.
PR 15-OCT-1993; GB-021301.
PR 15-OCT-1993; GB-021303.
PR 15-OCT-1993; GB-021305.
PR 15-OCT-1993; GB-021304.
PA (DANI-) DANISCO AS.
PI Bojko M, Bojlsen K, Christensen TMIE, Kragh KM, Marcussen J;
PI Nielsen J, Yu S;
DR WPI: 95-161801/21.
DR N-PSDB: Q87616.
PT 1,5-D-anhydrofructose prodn. from alpha-1,4-glucan - by treatment
PT with pure alpha-1,4-glucan lyase, partic. useful as antioxidant and
PT sweetener for foods and beverages
PS Claim 9; Page 95-99; 166pp; English.
CC An alpha-1,4-glucan lyase enzyme (preferably pullanase or isoamylase)
CC is used in a new method for the production of 1,5-D-anhydrofructose.

CC The enzyme is isolated from either a fungus (M.costata or M.vulgaris)
CC or from fungally infected algae (G.lemaleiformis) or algae alone.
CC 1,5-D-anhydrofructose is useful as an antioxidant and sweetener for
CC foodstuffs and beverages. It is also useful as an intermediate for the
CC antibiotic microthecin, an oxygen scavenger during polymerisation
CC reactions and as a reducing agent in the synthesis of biodegradable
CC plastics.
SQ Sequence 1066 AA;

Query Match 54.6%; Score 59; DB 14; Length 1066;
Best Local Similarity 41.7%; Pred. No. 3.26e+01;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 891 wtgggganrikf 902
| | : | : | : |
QY 26 WKGHVGQDRVDF 37

RESULT 10
ID R72711 standard; protein; 1066 AA.
AC R72711;
DT 27-DEC-1995 (revised)
DT 31-OCT-1995 (first entry)
DE Alpha-1,4-glucan lyase.
KW Alpha-1,4-glucan lyase; enzyme.
OS Morchella costata.
FH Key Location/Qualifiers
FT peptide 12..24
FT peptide 147..156 /note= "sequenced peptide"
FT peptide 161..280 /note= "sequenced peptide"
FT peptide 307..339 /note= "see above"
FT peptide 342..370 /note= "see above"
FT peptide 415..440 /note= "see above"
FT peptide 813..854 /note= "see above"
FT peptide 860..889 /note= "see above"
FT peptide 1001..1023 /note= "see above"
FT peptide 1001..1023 /note= "see above"
PN WO9510617-A.
PD 20-APR-1995.
PF 15-OCT-1994; E03398.
PR 15-OCT-1993; GB-021302.
PA (DANI-) DANISCO AS.
PI Bojlsen K, Christensen TMIE, Kragh KM, Marcussen J;
PI Yu S;
DR WPI: 95-161802/21.
DR N-PSDB: Q89701.
PT Isolation of alpha-1,4-glucan lyase from fungus - and its
PT recombinant prodn from isolated DNA.
PS Claim 6; Fig 7; 51pp; English.
CC The enzyme alpha-1,4-glucan lyase(GL) was purified from Morchella
CC costata (Mc) by affinity chromatography on beta-cyclodextrin
CC Sepharose (RTM), ion exchange on Mono Q HR 5/5 (RTM) and gel
CC filtration on Sepharose (RTM) 12 columns. The purified enzyme
CC appeared colourless. The mol. mass was 110 kDa (SDS-PAGE). It showed
CC an isoelectric point of pI 5.4. The optimum pH range for the fungal
CC lyase catalysed reaction was between pH 5 and pH 7. The purified
CC fungal lyase degraded maltosaccharides. The highest activity was
CC with maltotetraose (activity 100%), then maltohexaose (97%),
CC maltotriose (56%) and maltose (2%). The lyase was digested with
CC endoproteinase Arg-C from Clostridium histolyticum or endoproteinase
CC Lys-C from Lysobacter enzymes. The resulting peptides were
CC sequenced (see FT). Amino acid sequences of three overlapping
CC tryptic peptides from GL (see R72713) were used to generate mixed
CC oligos which could be used as PCR primers for the amplification of
CC DNA isolated from Morchella. The primers are Q89703-04 and Q90305-06.

```
SQ Sequence 1066 AA;
Query Match 54.6%; Score 59; DB 13; Length 1066;
Best Local Similarity 41.7%; Pred.No. 3.26e+01;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 891 wtggggnrikf 902
   | | | | |
QY 26 WKGHVGQDRVDF 37

RESULT 11
ID R79930 standard; Protein: 419 AA.
AC R79930;
DT 09-MAY-1996 (first entry)
DE Porcine acylglucosamine-2-epimerase mutant.
KW Porcine; acylglucosamine-2-epimerase; N-acetylmannosamine;
KW N-acetylneuraminic acid; renin-binding; enzymatic production;
KW mutant.
OS Sus scrofa.
FH Key Location/Qualifiers
FT misc_difference 10 /note= "wild type Ala substd. with Val"
FT misc_difference 21 /note= "wild type Arg substd. with Ser"
FT misc_difference 23 /note= "wild type Met substd. with Ile"
FT misc_difference 27 /note= "wild type Leu substd. with Met"
FT misc_difference 33 /note= "wild type Arg substd. with Gln"
FT misc_difference 47 /note= "wild type Arg substd. with Gln"
FT misc_difference 51 /note= "wild type Asp substd. with His"
FT misc_difference 71 /note= "wild type Lys substd. with Thr"
FT misc_difference 72 /note= "wild type Leu substd. with Phe"
FT misc_difference 76 /note= "wild type His substd. with Arg"
FT misc_difference 78 /note= "wild type Pro substd. with Val"
FT misc_difference 93 /note= "wild type Arg substd. with Ser"
FT misc_difference 94 /note= "wild type His Substd. with Tyr"
FT misc_difference 101 /note= "wild type Glu substd. with Gly"
FT misc_difference 110 /note= "wild type Arg substd. with Gln"
FT misc_difference 120 /note= "wild type Ser substd. with Thr"
FT misc_difference 136 /note= "wild type Arg substd. with Lys"
FT misc_difference 139 /note= "wild type Ala substd. with Gly"
FT misc_difference 141 /note= "wild type Ala substd. with Met"
FT misc_difference 142 /note= "wild type Arg substd. with His"
FT misc_difference 145 /note= "wild type Ser substd. with Arg"
FT misc_difference 149 /note= "wild type Asp substd. with Glu"
FT misc_difference 155 /note= "wild type Val substd. with Ile"
FT misc_difference 163 /note= "wild type Ser substd. with Ala"
FT misc_difference 171 /note= "wild type Pro substd. with Ser"
FT misc_difference 173 /note= "wild type Ala substd. with Thr"
FT misc_difference 174 /note= "wild type Val substd. with Leu"
FT misc_difference 176 /note= "wild type Ser substd. with Thr"
FT misc_difference 178 /note= "wild type Ser substd. with Pro"
FT misc_difference 187 /note= "wild type Cys substd. with Asn"
FT misc_difference 199 /note= "wild type Leu substd. with Met"
FT misc_difference 200 /note= "wild type Ala substd. with Thr"
FT misc_difference 201 /note= "wild type Gly substd. with Asp"
FT misc_difference 202 /note= "wild type Arg substd. with Lys"
FT misc_difference 208 /note= "wild type His substd. with Asp"
FT misc_difference 212 /note= "wild type Arg substd. with His"
FT misc_difference 224 /note= "wild type Ala substd. with Val"
FT misc_difference 234 /note= "wild type Glu substd. with Lys"
FT misc_difference 237 /note= "wild type Ser substd. with Pro"
FT misc_difference 249 /note= "wild type Ala substd. with Thr"
FT misc_difference 258 /note= "wild type Arg substd. with Gln"
FT misc_difference 259 /note= "wild type His substd. with Tyr"
FT misc_difference 260 /note= "wild type Ser substd. with Ala"
FT misc_difference 261 /note= "wild type Ser substd. with Leu"
FT misc_difference 263 /note= "wild type Ser substd. with Lys"
FT misc_difference 266 /note= "wild type Ala substd. with Pro"
FT misc_difference 269 /note= "wild type Arg substd. with Gln"
FT misc_difference 270 /note= "wild type Ala substd. with Arg"
FT misc_difference 272 /note= "wild type Val substd. with Ile"
FT misc_difference 275 /note= "wild type Thr substd. with Lys"
FT misc_difference 282 /note= "wild type Arg substd. with His"
FT misc_difference 287 /note= "wild type Ala substd. with Pro"
FT misc_difference 288 /note= "wild type Asp substd. with Glu"
FT misc_difference 300 /note= "wild type Gly substd. with Asp"
FT misc_difference 317 /note= "wild type Ser substd. with Thr"
FT misc_difference 328 /note= "wild type Ser substd. with Arg"
FT misc_difference 329 /note= "wild type Glu substd. with Asp"
FT misc_difference 337 /note= "wild type Arg substd. with Asn"
FT misc_difference 348 /note= "wild type Arg substd. with His"
FT misc_difference 364 /note= "wild type Arg substd. with Gln"
FT misc_difference 393 /note= "wild type Met substd. with Ile"
FT misc_difference 395 /note= "wild type Ser substd. with Gly"
FT misc_difference 399
```

```

FT      /note= "wild type Ser substd. with Gln"
FT      misc_difference 402
FT      /note= "wild type Ala substd. with Gly"
FT      misc_difference 403..419
FT      /note= "C-terminal addition to the wild type
FT      sequence"
FN      WO9526399-A1.
PD      05-OCT-1995.
PD      24-MAR-1995; J00541.
PR      25-MAR-1994; JP-056271.
PR      09-SEP-1994; JP-218333.
PA      (MARU ) MARUKIN SHOU KK.
PI      Maru I, Ohta Y, Tsukada Y;
PI      WPI; 95-351320/45.
PT      Recombinant acyl:glucosamine-2-epimerase with renin-binding activity
PT      - useful in enzymic production of N-acetylmannosamine and
PT      N-acetylneuraminic acid
PS      Claim 13; Pages 61-63; 74pp; Japanese.
CC      R79930 is a porcine acylglucosamine-2-epimerase (A2P) mutant,
CC      with renin-binding activity. A2P can be used for the enzymatic
CC      prodn. of N-acetylmannosamine and N-acetylneuraminic acid.
SQ      Sequence 419 AA;

      Query Match      52.8%; Score 57; DB 16; Length 419;
      Best Local Similarity 66.7%; Pred. NO. 5.28e+01;
      Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      10 vvkqrvge 18
      ||| :|||:
Qy      25 VMKHVGQD 33

RESULT 12
ID      W36499 standard; Protein; 348 AA.
AC      W36499.
DT      14-APR-1998 (first entry)
DE      Hereditary haemochromatosis gene product.
KW      Hereditary haemochromatosis; metal toxicity; diagnosis;
KW      gene therapy; prenatal screening; human.
OS      Homo sapiens.
FH      Key Location/Qualifiers
FT      Misc_difference 63
FT      /note= "substituted by Asp in 24s2 mutant"
FT      Misc_difference 65
FT      /note= "substituted by Cys in 24d7 variant"
FT      Misc_difference 282
FT      /note= "substituted by Tyr in 24d1 mutant"
PN      WO9738137-A1.
PD      16-OCT-1997.
PF      04-APR-1997; U06254.
PR      23-MAY-1996; US-652265.
PR      04-APR-1996; US-630912.
PR      16-APR-1996; US-632673.
PA      (MERC-) MERCATOR GENETICS INC.
PI      Drayna DT, Feder JN, Gairke A, Ruddy D, Thomas WJ,
PI      Tsuchihashi Z, Wolff RK;
PI      WPI; 97-512743/47.
DR      N-PSDB; T96690-91.
PT      Hereditary haemochromatosis gene and variants - useful for diagnosis
PT      and treatment of hereditary haemochromatosis disease
PS      Disclosure; Fig 4; 115pp; English.
CC      This polypeptide is the expression product of a novel human gene
CC      (see T96650) whose mutated form is associated with hereditary
CC      haemochromatosis (HH). A single mutation (24d1) in the HH gene
CC      appears responsible for the majority of HH disease. This comprises
CC      a G to A substitution that is present in 86% of affected
CC      chromosomes and in 4% of unaffected chromosomes. It results in a
CC      Cys to Tyr substitution in the encoded protein at a critical
CC      disulphide bridge important for secondary structure. The following
CC      are claimed: the 10825 bp genomic DNA sequence (1), a 1437 bp cDNA
CC      sequence (1a) (see T96691) and their 24d1, 24d2 and 24d7 variants;
CC      a cloning or expression vector; host cells; a peptide product
CC      chosen from the HH gene product, its variants (24d1, 24d2 and

```

```

CC      24d7), or a peptide of at least 56 amino acid residues of these; an
CC      antibody produced using the peptide as an immunogen; a method to
CC      determine the presence or absence of the common HH gene mutation;
CC      an animal model for the HH disease; metal chelation agents, T-cell
CC      differentiation factors and therapeutic agents for the mitigation
CC      of injury due to oxidative process in vivo or mitigation of iron
CC      overload; a method for screening potential therapeutic agents for
CC      activity in connection with HH disease; an antisense oligonucleotide
CC      directed against a transcriptional product of a nucleic acid
CC      sequence as above; and oligonucleotides or pairs of oligonucleotides
CC      covering a range of nucleotides from (1), (1a) or their variants,
CC      useful for detecting a polymorphism in the HH gene. The invention
CC      also relates to methods for screening for HH homozygotes, to HH
CC      diagnosis, prenatal screening and diagnosis, and therapies of HH
CC      disease, including gene therapy, protein- and antibody-based
CC      therapeutics, and small molecule therapeutics.
SQ      Sequence 348 AA;

```

```

      Query Match      51.9%; Score 56; DB 27; Length 348;
      Best Local Similarity 53.8%; Pred. NO. 6.70e+01;
      Matches 7; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

```

```

Db      136 wkygydsgdhllef 148
      ||| :|||:|
Qy      26 WK-GHVGQDRVDF 37

```

```

RESULT 13
ID      W52352 standard; Protein; 291 AA.
AC      W52352.
DT      09-JUL-1998 (first entry)
DE      Bacillus subtilis protopectinase.
KW      Protopectinase; protopectin degradation; fibre production; pectin;
KW      pulp.
OS      Bacillus subtilis.
PN      WO9806832-A1.
PD      19-FEB-1998.
PF      08-AUG-1997; J02794.
PR      12-AUG-1996; JP-212919.
PA      (SAKA/) SAKAI T.
PI      Sakai T;
DR      WPI; 98-159525/14.
DR      N-PSDB; V19947.
PT      Protopectinase - is useful in fibre production, and preparation of
PT      pectin and pulp
PS      Claim 4; Pages 16-17; 27pp; Japanese.
CC      The present sequence is Bacillus subtilis protopectinase,
CC      which degrades protopectin, has a molecular weight of 30000 as
CC      determined by SDS-PAGE, has an optimum pH of 6.0 and temperature of
CC      60 degrees C and is inhibited by Hg and Mn. The protopectinase can
CC      be used in fibre production, and in the preparation of pectin and
CC      pulp.
SQ      Sequence 291 AA;

```

```

      Query Match      50.9%; Score 55; DB 30; Length 291;
      Best Local Similarity 63.6%; Pred. NO. 8.50e+01;
      Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

Db      248 wkgpgggqdivn 258
      ||| :|||:|
Qy      26 WKGHVGQDRVD 36

```

```

RESULT 14
ID      W49875 standard; Protein; 319 AA.
AC      W49875.
DT      21-DEC-1998 (first entry)
DE      Pyrococcus VC1-7EG1 glycosidase.
KW      Glycosidase; VC1-7EG1; thermostable enzyme; oligosaccharide;
KW      glucose; sugar; baking; textile; detergent.
OS      Pyrococcus furiosus strain VC1-7EG1.
PN      WO9824799-A1.
PD      11-JUN-1998.

```

PF 08-DEC-1997; U22623.
PR 10-OCT-1997; US-949026.
PR 06-DEC-1996; US-056916.
PA (DIVE-) DIVERSA CORP.
PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
DR WPI: 98-362407/31.
DR N-PSDB; V36924.
PT Glycosidase enzymes from organisms of the genera Staphylothermus,
PT Pyrococcus and Thermococcus - for deriving sugar from
PT oligosaccharides, useful in the e.g. food processing, textile or
PT baking industries
PS Claim 1; Fig 18a-b; 92pp; English.
CC This is the amino acid sequence of glycosidase VC1-7EG1, deduced
CC from a polynucleotide (see V36924) of a clone (7EG1) of Pyrococcus
CC furiosus VC1, which grows optimally at 100 degC. The invention
CC provides 18 polynucleotides (see V36907-24) coding for thermostable
CC glycosidases (see W49858-75) having glucosidase, alpha-galactosidase,
CC beta-galactosidase, beta-mannosidase, beta-mannanase, endoglucanase
CC or pullulanase activity. Vectors and host cells are also claimed.
CC A method is provided for producing the enzymes by recombinant
CC techniques. A claimed method for generating glucose from soluble
CC cell oligosaccharides comprises contacting a sample (selected from
CC dairy products, fruit juice, detergent, textile, guar gum, animal
CC feed, plant biomass or waste product) containing oligosaccharides
CC (selected from maltose, cellobiose, lactose, sucrose, raffinose,
CC stachyose, verbascose, cellulose, starch, amylose, glycogen,
CC disaccharides, polysaccharides and pullulan) with one of the
CC claimed glycosidases such that glucose is produced.
SQ Sequence 319 AA;

Query Match 50.9%; Score 55; DB 36; Length 319;
Best Local Similarity 46.2%; Pred. No. 8.50e+01;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 234 vkkanigweyvaf 246
QY 25 VMKGHVQDRVDF 37
|||::|:|

Search completed: Thu Jul 8 18:10:03 1999
Job time : 24 secs.

Query Match 50.9%; Score 55; DB 36; Length 319;
Best Local Similarity 46.2%; Pred. No. 8.50e+01;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 234 vkkanigweyvaf 246
QY 25 VMKGHVQDRVDF 37
|||::|:|

RESULT 15
ID W35007 standard; Protein; 319 AA.
AC W35007;
DT 21-MAY-1998 (first entry)
DE Pyrococcus furiosus endoglucanase.
KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase.
OS Pyrococcus furiosus (Clone 7EG1).
PN W09744361-A1.
PD 27-NOV-1997.
PF 22-MAY-1997; U08793.
PR 22-MAY-1996; US-651572.
PA (RECO-) RECOMBINANT BIOTOCALYSIS INC.
PI Lam DE, Mathur EJ;
DR WPI: 98-018435/02.
DR N-PSDB; T94215.
PT Endoglucanase(s), preferably from archaeal bacterium, AEPII la -
PT useful to degrade carboxymethylcellulose and hydrolyse of
PT beta-1,4-glycosidic bonds in cellulose
PS Claim 1; Fig 1W; 164pp; English.
CC This protein comprises an endoglucanase of Pyrococcus furiosus
CC (Clone 7EG1) that is capable of degrading carboxymethylcellulose
CC and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It
CC has homology to an endoglucanase of archaeobacterium AEPIIa (see
CC W34985). It can be produced from native cells or from recombinant
CC host cells, especially prokaryotic host cells transformed with a
CC plasmid or virus-derived vector including the endoglucanase DNA
CC (see 194215). 24 Endoglucanases (see W34986-W35008) are claimed.
CC They can be used to degrade cellulose for the conversion of plant
CC biomass into fuels and chemicals, for use in detergents, textiles,
CC animal feed, waste treatment, and in the fruit juice and brewing
CC industries for the clarification and extraction of juices.
SQ Sequence 319 AA;

THIS PAGE BLANK (USPTO)

W P E R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run On: Thu Jul 8 18:09:07 1999; MasPar time 3.49 Seconds
Tabular output not generated. 149.177 Million cell updates/sec

Title: >US-09-041-236-2
Description: (25-37) from US09041236.pep (3 of 45)
Perfect Score: 108
Sequence: 1 VNKGHVQDRVDF 13

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 26.454; Variance 36.593; scale 0.723

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	65	60.2	403	2	JC5269 neuron-specific signa	1.68e-01
2	63	58.3	402	2	JC5270 neuron-specific signa	4.12e-01
3	61	56.5	167	2	G69972 hypothetical protein	9.93e-01
4	61	56.5	894	2	B42372 sensor protein kdpd	9.93e-01
5	59	54.6	1308	2	B32494 transposable element	2.35e+00
6	58	53.7	925	2	G69539 hypothetical protein	3.60e+00
7	57	52.8	250	2	A70579 probable yfiH protein	5.48e+00
8	57	52.8	254	2	S77113 hypothetical protein	5.48e+00
9	57	52.8	347	2	S50403 TIF34 protein - yeast	5.48e+00
10	57	52.8	353	2	H69418 hypothetical protein	5.48e+00
11	57	52.8	419	2	JX0187 renin-binding protein	5.48e+00
12	57	52.8	595	2	S72537 acr-2 protein - Neuro	5.48e+00
13	57	52.8	595	2	S78458 acr-2 protein - Neuro	5.48e+00
14	57	52.8	1209	2	T00373 KIAA0649 protein - hu	5.48e+00
15	56	51.9	472	1	WVADF6 early E1B 52K protein	8.29e+00
16	56	51.9	1321	2	T00382 hypothetical protein	8.29e+00
17	56	51.9	2255	1	J01532 genome polyprotein -	8.29e+00
18	55	50.9	320	2	S36092 pyridoxamine-phosphat	1.25e+01
19	55	50.9	313	2	E59580 arabinan-endo 1,5-alp	1.25e+01
20	55	50.9	322	2	S51376 sucrose cleavage prot	1.25e+01
21	55	50.9	707	2	S29854 spectrin beta chain -	1.25e+01
22	55	50.9	837	1	A31842 endo-1,4-beta-xylanas	1.25e+01
23	55	50.9	2364	2	A44159 spectrin beta-G chain	1.25e+01

24	54	50.0	165	2	S44302 single-stranded DNA-b	1.87e+01
25	54	50.0	293	2	B70455 thiosulfate sulfurtra	1.87e+01
26	54	50.0	317	3	T00146 hypothetical protein	1.87e+01
27	54	50.0	388	2	S42122 hypothetical protein	1.87e+01
28	54	50.0	469	2	S74825 hypothetical protein	1.87e+01
29	54	50.0	1155	2	G64332 FUN12/bif-2 family pr	1.87e+01
30	53	49.1	183	1	D64430 probable transcriptio	2.78e+01
31	53	49.1	230	2	C71548 probable ABC transpor	2.78e+01
32	53	49.1	260	2	C65061 hypothetical protein	2.78e+01
33	53	49.1	300	2	A70819 hypothetical protein	2.78e+01
34	53	49.1	349	2	S47093 hypothetical protein	2.78e+01
35	53	49.1	359	2	JC5382 hereditary hemochroma	2.78e+01
36	53	49.1	607	2	S52629 catechol oxidase (EC	2.78e+01
37	53	49.1	1110	2	I59370 guanylate cyclase (EC	2.78e+01
38	52	48.1	145	2	S34574 gene 70 protein - Myc	4.12e+01
39	52	48.1	290	2	H64434 ribosomal protein S6	4.12e+01
40	52	48.1	291	2	D64377 ribosomal protein S6	4.12e+01
41	52	48.1	402	2	A35741 renin-binding protein	4.12e+01
42	52	48.1	415	2	C71467 probable tyrosine tra	4.12e+01
43	52	48.1	417	2	JX0188 renin-binding protein	4.12e+01
44	52	48.1	943	2	F69543 ATP-dependent RNA hel	4.12e+01
45	52	48.1	3085	3	T00327 polyprotein - infecti	4.12e+01

ALIGNMENTS

RESULT 1
ENTRY JC5269 #type complete
TITLE neuron-specific signal transduction protein Stac - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 25-Apr-1997 #sequence_revision 09-May-1997 #text_change 01-May-1998
ACCESSIONS JC5269
REFERENCE JC5269
#authors Suzuki, H.; Kawai, J.; Taga, C.; Yaol, T.; Hara, A.; Hirose, K.; Hayashizaki, Y.; Watanabe, S.
#journal Biochem. Biophys. Res. Commun. (1996) 229:902-909
#title Stac, a novel neuron-specific protein with cysteine-rich and SH3 domains.
#cross-references MUID:97115677
#status JC5269
#nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-403 #label SUZ
#cross-references DBJ:D86639; NID:g1799565; PID:d1013839; PID:g1799566
COMMENT This protein has a Src homology 3 and a cysteine-rich domains. It is involved in a neuron-specific signal transduction.

GENETICS
#gene Stac
CLASSIFICATION #superfamily protein kinase C zinc-binding repeat homology; SH3 homology
FEATURE 109-160 #domain protein kinase C zinc-binding repeat homology #label KZ2
293-340 #domain SH3 homology #label SH3
SUMMARY #length 403 #molecular-weight 44319 #checksum 1692
Query Match 60.2%; Score 65; DB 2; Length 403;
Best Local Similarity 66.7%; Pred. No. 1.68e-01;
Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;
Db 324 WKGKI-QDRVGF 334
||| : |||||
Qy 26 WKGHVQDRVDF 37
RESULT 2
ENTRY JC5270 #type complete
TITLE neuron-specific signal transduction protein Stac - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 25-Apr-1997 #sequence_revision 09-May-1997 #text_change 01-May-1998
ACCESSIONS JC5270

```

REFERENCE
#authors      JC5269
#journal      Suzuki, H.; Kawai, J.; Taga, C.; Yaol, T.; Hara, A.; Hirose,
#title       K.; Hayashizaki, Y.; Watanabe, S.
#journal      Biochem. Biophys. Res. Commun. (1996) 229:902-909
#title       Stac, a novel neuron-specific protein with cysteine-rich and
#journal      SH3 domains.
#cross-references MUID:97115677
#accession      JC5270
#status        nucleic acid sequence not shown
#molecule_type mRNA
#residues       1-402 #label SUZ
#cross-references DBJ:D86640; NID:gl799567; PID:d1013840; PID:gl799568
COMMENT        This protein has a Src homology 3 and a cysteine-rich domains. It
               is involved in a neuron-specific signal transduction.

GENETICS
#gene          Stac
CLASSIFICATION #superfamily protein kinase C zinc-binding repeat homology;
               SH3 homology
FEATURE
108-159        #domain protein kinase C zinc-binding repeat homology
               #label K22\
292-339        #domain SH3 homology #label SH3
               #length 402 #molecular-weight 44553 #checksum 1906
SUMMARY
Query Match    58.3%; Score 63; DB 2; Length 402;
Best Local Similarity 58.3%; Pred. No. 4.12e-01;
Matches        7; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Db 323 WKGI-QDRIGF 333
| | | : | | | : |
QY 26 WKGHVGQDRVDF 37

RESULT 3
ENTRY 3
TITLE hypothetical protein yrdA - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
G69972
A69580
#authors      Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
               Alloni, G.; Azevedo, V.; Bextero, M.G.; Bessieres, P.;
               Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
               A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
               Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
               Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
               Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
               Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
               Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
               M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
               S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
               Guisepi, G.; Guy, B.J.; Haga, K.; Haelech, J.; Harwood,
               C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
               Huilo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
               Kasahara, Y.; Klaier-Blanchard, M.; Klein, C.; Kobayashi,
               Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
               Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
               Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
               Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
               M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
               M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
               V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
               A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
               Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
               Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
               Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
               Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
               B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
               Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
               Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
               Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
               Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;

```

```

Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
#journal      Nature (1997) 390:249-256
#title       The complete genome sequence of the Gram-positive bacterium
#journal      Bacillus subtilis.
#cross-references MUID:98044033
#accession      G69972
#status        preliminary; nucleic acid sequence not shown;
               translation not shown
#molecule_type DNA
#residues       1-167 #label KUN
#cross-references GB:J99117; GB:AL009126; NID:g2634966; PID:el183907;
               PID:g2635123
#experimental_source strain 168

GENETICS
#gene          yrdA
SUMMARY        #length 167 #molecular-weight 18917 #checksum 37
Query Match    56.5%; Score 61; DB 2; Length 167;
Best Local Similarity 41.7%; Pred. No. 9.93e-01;
Matches        5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 99 WTDHTLQERINF 110
| | | : | | | : |
QY 26 WKGHVGQDRVDF 37

RESULT 4
ENTRY 4
TITLE sensor protein kdpB (EC 2.7.3.-) - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 10-Jul-1992 #sequence_revision 10-Jul-1992 #text_change
17-Jul-1998
B42372
A42372
#authors      Walderhaug, M.O.; Polarek, J.W.; Voelkner, P.; Daniel, J.M.;
               Hesse, J.E.; Altendorf, K.; Epstein, W.
               J. Bacteriol. (1992) 174:2152-2159
               KdpB and KdpE, proteins that control expression of the kdpABC
               operon, are members of the two-component sensor-effector
               class of regulators.
#cross-references MUID:92202141
#accession      B42372
#status        preliminary
#molecule_type DNA
#residues       1-894 #label WAL
#cross-references GB:M36086; NID:g146550; PID:g146551
A64720
#authors      Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
               Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
               Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
               Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
               Y.
               Science (1997) 277:1453-1462
#journal      The complete genome sequence of Escherichia coli K-12.
#title       #cross-references MUID:97426617
#accession      F64804
#status        nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues       1-894 #label BLAT
#cross-references GB:AE000173; GB:U000096; NID:gl786910; PID:gl786912;
               UWPG:b0695
#experimental_source strain K-12, substrain MG1655

GENETICS
#gene          kdpB
CLASSIFICATION #superfamily sensor histidine kinase homology
KEYWORDS        autophosphorylation; P-loop; phosphohistidine;
               phosphoprotein; phosphotransferase; transmembrane protein
FEATURE
31-38          #region nucleotide-binding motif A (P-loop)\
401-417        #domain transmembrane #status predicted #label TM1\
426-442        #domain transmembrane #status predicted #label TM2\

```

```

449-465 #domain transmembrane #status predicted #label TM3\
478-494 #domain transmembrane #status predicted #label TM4\
643-880 #domain sensor histidine kinase homology #label SHK\
840-856 #domain transmembrane #status predicted #label TM5\
673 #binding_site phosphate (His) (covalent) (by
autophosphorylation) #status predicted
SUMMARY #length 894 #molecular-weight 98717 #checksum 4705

Query Match 56.5%; Score 61; DB 2; Length 894;
Best Local Similarity 50.0%; Pred. No. 9.92e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 237 WRGHPGEEKV 246
QY 26 WKGHVGQDRV 35
I::I I::I

RESULT 5
ENTRY B32494 #type complete
TITLE transposable element TxlC protein 2 - African clawed frog
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change
10-Sep-1997

ACCESSIONS B32494
REFERENCE A32494
#authors Garrett, J.E.; Knutzon, D.S.; Carroll, D.
#journal Mol. Cell. Biol. (1989) 9:3018-3027
#title Composite transposable elements in the Xenopus laevis genome.
#cross-references MUID:89384562
#accession B32494
##status preliminary
##molecule_type DNA
##residues 1-1308 #label GAR
##cross-references GB:M26915; NID:g214844; PID:g214846
SUMMARY #length 1308 #molecular-weight 149576 #checksum 9519

Query Match 54.6%; Score 59; DB 2; Length 1308;
Best Local Similarity 58.3%; Pred. No. 2.35e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 194 VRDGHVSQSRID 205
QY 25 WKGHVGQDRV 36
I::I I::I

RESULT 6
ENTRY S69539 #type complete
TITLE hypothetical protein 31 - phage HP1
ORGANISM #formal_name Phage HP1
DATE 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change
17-Mar-1999

ACCESSIONS S69539
REFERENCE S69503
#authors Esposito, D.; Fitzmaurice, W.P.; Benjamin, R.C.; Goodman,
S.D.; Waldman, A.S.; Scoocca, J.J.
#journal Nucleic Acids Res. (1996) 24:2360-2368
#title The complete nucleotide sequence of bacteriophage HP1 DNA.
#cross-references MUID:96279738
#accession S69539
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-925 #label ESP
##cross-references EMBL:U241159; NID:gi046235; PID:gi046260
##note the nucleotide sequence was submitted to the EMBL Data
Library, April 1995
SUMMARY #length 925 #molecular-weight 102000 #checksum 8473

Query Match 53.7%; Score 58; DB 2; Length 925;
Best Local Similarity 38.5%; Pred. No. 3.60e+00;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 813 WQGNVSSGRINI 825

```

```

QY 25 WKGHVGQDRVDF 37
I::I I::I

RESULT 7
ENTRY A70579 #type complete
TITLE probable yfiH protein - Mycobacterium tuberculosis (strain
H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS A70579
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession A70579
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-250 #label COL
##cross-references GB:Z95388; GB:AL123456; NID:g3261759; PID:e315955;
PID:g2104329
##experimental_source strain H37Rv
GENETICS
#gene yfiH
#summary #length 250 #molecular-weight 25964 #checksum 8371
Query Match 52.8%; Score 57; DB 2; Length 250;
Best Local Similarity 46.2%; Pred. No. 5.48e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 71 VWMNQVHGDRVEL 83
QY 25 WKGHVGQDRVDF 37
I::I I::I

RESULT 8
ENTRY S77113 #type complete
TITLE hypothetical protein sll1771 - Synechocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synechocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S77113
REFERENCE S74322
#authors Kaneo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synechocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S77113
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA

```

```
##residues      1-254 ##label KAN
##cross-references EMBL:D90908; GB:AB001339; NID:g1552725; PID:d1018404;
                  PID:g1652752
##note
the nucleotide sequence was submitted to the EMBL Data
Library, June 1996

GENETICS
#start_codon   GTG
CLASSIFICATION #superfamily conserved hypothetical protein yloO: conserved
                hypothetical protein yloO homology
FEATURE
40-252        #domain conserved hypothetical protein yloO homology
                #label yloO
SUMMARY
#length 254 #molecular-weight 28472 #checksum 4749
Query Match      52.8%; Score 57; DB 2; Length 254;
Best Local Similarity 50.0%; Pred. No. 5.48e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 119 WRAHVGD SRL 128
QY 26 WKGHVGQDRV 35
      |::|::|:|:|

RESULT 9
ENTRY S50403 #type complete
TITLE Tif34 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YN9375.16c; protein YMR146c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 13-Jan-1995 #sequence_revision 24-Feb-1995 #text_change
13-Sep-1998
ACCESSIONS S50403
REFERENCE S50388
#authors Badcock, K.; Churcher, C.
#submission submitted to the EMBL Data Library, December 1994
#accession S50403
#molecule_type DNA
##residues 1-347 ##label BAD
##cross-references EMBL:Z47071; NID:g606429; PID:g606445; MIPS:YMR146c
#gene SGD:TIF34
##cross-references SGD:S0004754; MIPS:YMR146c
#map_position 13R
CLASSIFICATION #superfamily TGF-beta receptor interacting protein; WD repeat
                homology
SUMMARY #length 347 #molecular-weight 38755 #checksum 4129
Query Match      52.8%; Score 57; DB 2; Length 347;
Best Local Similarity 50.0%; Pred. No. 5.48e+00;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 90 WKSPPVPVKRVEF 101
QY 26 WKGHVGQDRVDF 37
      |::|::|:|:|

RESULT 10
ENTRY H69418 #type complete
TITLE hypothetical protein AF1353 - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Jun-1998
ACCESSIONS H69418
REFERENCE A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kierlavage, A.R.; Graham,
D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
```

```
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
Nature (1997) 390:364-370
The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession H69418
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-353 ##label KLE
#cross-references GB:AE001010; GB:AE000782; NID:g2689333; PID:g2649225;
TIGR:AF1353
SUMMARY #length 353 #molecular-weight 39283 #checksum 2579
Query Match      52.8%; Score 57; DB 2; Length 353;
Best Local Similarity 55.6%; Pred. No. 5.48e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 119 WNGEYGEER 127
QY 26 WKGHVGQDR 34
      |::|::|:|:|

RESULT 11
ENTRY JX0187 #type complete
TITLE renin-binding protein - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
20-Mar-1998
ACCESSIONS JX0187
REFERENCE JX0187
#authors Inoue, H.; Takahashi, S.; Fukui, K.; Miyake, Y.
#journal J. Biochem. (1991) 110:493-500
#title Genetic and molecular properties of human and rat
                renin-binding proteins with reference to the function of
                the leucine zipper motif.
#cross-references MUID:92138649
#accession JX0187
#molecule_type mRNA
#residues 1-419 ##label INO
#cross-references GB:D10233; GB:D01086; NID:g220896; PID:d1001552;
PID:g220897
COMMENT Renin-binding protein is a protein that binds to renin to form a
                high molecular weight renin.
KEYWORDS leucine zipper
FEATURE #region hydrophobic\
184-189 #region leucine zipper motif
185-206
SUMMARY #length 419 #molecular-weight 48391 #checksum 3024
Query Match      52.8%; Score 57; DB 2; Length 419;
Best Local Similarity 66.7%; Pred. No. 5.48e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 10 VMKQRVGQE 18
QY 25 VMKGHVGQD 33
      |::|::|:|:|

RESULT 12
ENTRY S72537 #type complete
TITLE acr-2 protein - Neurospora crassa
ORGANISM #formal_name Neurospora crassa
DATE 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change
02-Jul-1998
ACCESSIONS S72537
REFERENCE S72535
#authors Akiyama, M.; Nakashima, H.
#journal Biochim. Biophys. Acta (1996) 1307:187-192
#title Molecular cloning of the acr-2 gene which controls
                acriflavine sensitivity in Neurospora crassa.
```

```
#cross-references MUID:96283814
#accession S72537
#status preliminary
#molecule_type DNA
##residues 1-595 #label AKI
##cross-references EMBL:D45893
GENETICS
#gene acr-2
#introns 203/3
CLASSIFICATION #superfamily GAL4 zinc binuclear cluster homology
FEATURE
17-54
SUMMARY
#domain GAL4 zinc binuclear cluster homology #label GAL4
#length 595 #molecular-weight 65325 #checksum 7540
Query Match 52.8%; Score 57; DB 2; Length 595;
Best Local Similarity 66.7%; Pred. No. 5.48e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 314 WKAHVGAAR 322
QY 26 WKGHVGQDR 34
|||||
RESULT 13
ENTRY S78458 #type complete
TITLE acr-2 protein - Neurospora crassa
ORGANISM #formal_name Neurospora crassa
DATE 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change
02-Jul-1998
ACCESSIONS S78458
REFERENCE S78458
#authors Akiyama, M.
#submission submitted to the EMBL Data Library, February 1995
#accession S78458
#status preliminary
#molecule_type DNA
##residues 1-595 #label AKI
##cross-references EMBL:D45893; NID:g1754593; PID:g1754596
GENETICS
#gene acr-2
#introns 203/3
CLASSIFICATION #superfamily GAL4 zinc binuclear cluster homology
FEATURE
17-54
SUMMARY
#domain GAL4 zinc binuclear cluster homology #label GAL4
#length 595 #molecular-weight 65291 #checksum 8014
Query Match 52.8%; Score 57; DB 2; Length 595;
Best Local Similarity 66.7%; Pred. No. 5.48e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 314 WKAHVGAAR 322
QY 26 WKGHVGQDR 34
|||||
RESULT 14
ENTRY T00373 #type complete
TITLE KIAA0649 protein - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
01-Feb-1999
ACCESSIONS T00373
REFERENCE Z14142
#authors Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
#journal DNA Res. (1998) 5:169-176
#title Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.
#accession T00373
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
##residues 1-1209 #label ISH
```

```
#cross-references EMBL:AB014549; NID:d1204304; PID:d1032585
##experimental_source brain
GENETICS
#gene KIAA0649
SUMMARY
#length 1209 #molecular-weight 127351 #checksum 5514
Query Match 52.8%; Score 57; DB 2; Length 1209;
Best Local Similarity 60.0%; Pred. No. 5.48e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1050 VWRGGVGSR 1059
QY 25 VWRGHVGQDR 34
|||||
RESULT 15
ENTRY WNADE6 #type complete
TITLE early E1B 52K protein I - human adenovirus 41
ORGANISM #formal_name Mastadenovirus h41 #common_name human adenovirus 41
DATE 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
05-Sep-1997
ACCESSIONS F27333; B40249
REFERENCE A27333
#authors van Loon, A.E.; Ligtenberg, M.; Reemst, A.M.C.B.; Sussenbach, J.S.; Rozijn, T.H.
#journal Gene (1987) 58:109-126
#title Structure and organization of the left-terminal DNA regions of fastidious adenovirus types 40 and 41.
#cross-references MUID:88084437
#accession F27333
#molecule_type DNA
##residues 1-472 #label VAN
##cross-references GB:M18289; NID:g209778; PID:g209781
REFERENCE A40249
#authors Allard, A.; Wadell, G.
#journal Virology (1992) 188:319-330
#title The E1B transcription map of the enteric adenovirus type 41.
#cross-references MUID:92230230
#accession B40249
#molecule_type DNA
##residues 1-54; 57-233, 'P', 235-472 #label ALL
##cross-references GB:M87544
CLASSIFICATION #superfamily adenovirus early E1B protein I
KEYWORDS early protein
SUMMARY
#length 472 #molecular-weight 52157 #checksum 4733
Query Match 51.9%; Score 56; DB 1; Length 472;
Best Local Similarity 66.7%; Pred. No. 8.29e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 270 WKGLVGQNK 278
QY 26 WKGHVGQDR 34
|||||
Search completed: Thu Jul 8 18:09:19 1999
Job time : 12 secs.
```

(MT)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on: Thu Jul 8 18:08:00 1999; MasPar time 2.35 Seconds
156.427 Million cell updates/sec
Tabular output not generated.
```

```
>US-09-041-236-2
Description: (25-37) from US09041236.pep (3 of 45)
Perfect Score: 108
Sequence: 1 VWKGHVGQDRVDF 13
```

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database: swiss-prot37
          1:swissprot
```

Statistics: Mean 27.211; Variance 33.032; scale 0.824

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	61	56.5	187	YRDA_BACSU	HYPOTHETICAL 18.9 KD P	2.93e-01
2	61	56.5	894	KDPD_ECOLI	SENSOR PROTEIN KDPD (E	2.93e-01
3	59	54.6	1308	YTX2_XENLA	TRANSPOSON TX1 HYPOTHE	7.63e-01
4	58	53.7	925	VPH_BPHP1	PROBABLE TAIL FIBRE PR	1.12e+00
5	57	52.8	347	F34_YEAST	EUKARYOTIC TRANSLATION	1.94e+00
6	57	52.8	419	RNBP_RAT	RENIN-BINDING PROTEIN	1.94e+00
7	56	51.9	348	HFE_HUMAN	HEREDITARY HAEMOCHROMA	3.07e+00
8	56	51.9	360	HFE_RAT	HEREDITARY HAEMOCHROMA	3.07e+00
9	56	51.9	472	E1BL_ADB41	ELB PROTEIN, LARGE T-A	3.07e+00
10	56	51.9	2255	RRPL_SV5	RNA POLYMERASE BETA SU	3.07e+00
11	56	51.9	2255	RRPL_SV5WR	RNA POLYMERASE BETA SU	3.07e+00
12	55	50.9	270	PDHX_MYXXA	PROBABLE PYRIDOXAMINE	4.82e+00
13	55	50.9	837	YXN2_CLOTM	ENDO-1,4-BETA-XYLANASE	4.82e+00
14	55	50.9	2363	SPCO_MOUSE	SPECTRIN BETA CHAIN, B	4.82e+00
15	55	50.9	2364	SPCO_HUMAN	SPECTRIN BETA CHAIN, B	4.82e+00
16	54	50.0	165	SSB_PSEAE	SINGLE-STRAND BINDING	7.53e+00
17	54	50.0	302	CH12_GOSHI	ENDOCHITINASE 2 PRECUR	7.53e+00
18	54	50.0	324	CH1L_GOSHI	ENDOCHITINASE 1 PRECUR	7.53e+00
19	54	50.0	388	Y464_MYCCA	HYPOTHETICAL 44.5 KD P	7.53e+00
20	54	50.0	1155	IF2P_METJA	PROBABLE TRANSLATION I	7.53e+00
21	53	49.1	183	NUSA_METJA	NUSA PROTEIN HOMOLOG.	1.17e+01
22	53	48.1	260	YOCB_ECOLI	HYPOTHETICAL 29.7 KD P	1.17e+01
23	53	49.1	349	LB4D_RABIT	NADP-DEPENDENT LEUKOTR	1.17e+01

AC P21865;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SENSOR PROTEIN KDPD (EC 2.7.3.-).
GN KDPD.
OS ESCHERICHIA COLI.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
CC [1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12.
RX MEDLINE: 92202141.
RA WALTERHAUG M.O., POLAREK J.W., VOELKNER P., DANIEL J.M., HESSE J.E.,
RT ALTENDORF K., EPSTEIN W.;
RA "KdpD and KdpE, proteins that control expression of the kdpABC
RT operon, are members of the two-component sensor-effector class of
RT regulators.";
RL J. BACTERIOL. 174:2152-2159(1992).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., ROBE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL SCIENCE 277:1453-1474(1997).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-K12.
RX MEDLINE: 97061202.
RA OSHIMA T., AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,
RA IKEMOTO K., INADA T., ITOH T., KAJITHARA M., KANAI K., KASHIMOTO K.,
RA KIMURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K.,
RA MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SAITO N.,
RA SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,
RA YANO M., HORIUCHI T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA RES. 3:137-155(1996).
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM KDPD/KDPE
CC INVOLVED IN THE REGULATION OF THE KDP OPERON. KDPD MAY FUNCTION
CC AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT PHOSPHORYLATES KDPD
CC IN RESPONSE TO ENVIRONMENTAL SIGNALS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (PROBABLE).
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M36066; G146551; -
DR EMBL: AE000173; G1786912; -
DR EMBL: D90708; G1651302; -
DR PIR: B42372; B42372.
DR ECOGENE: EG10516; KDPD.
DR PFAM: PF00512; signal. 1.
KW SENSORY TRANSDUCTION; TRANSFERASE; KINASE; PHOSPHORYLATION;
KW TRANSMEMBRANE; INNER MEMBRANE.
FT TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 403 423 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT TRANSMEM 446 466 POTENTIAL.
FT TRANSMEM 478 498 POTENTIAL.
FT DOMAIN 650 894 TRANSMITTER DOMAIN (POTENTIAL).

FT MOD_RES 673 673 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 894 AA; 98718 MW; D1DD9FC7 CRC32;
Query Match 56.5%; Score 61; DB 1; Length 894;
Best Local Similarity 50.0%; Pred. No. 2.93e-01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 237 WRGHGEEKV 246
QY 26 WKGHVGQDRV 35
RESULT 3
ID YTX2_XENLA STANDARD; PRT; 1308 AA.
AC P14381;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DE TRANSPOSON TX1 HYPOTHETICAL 149 KD PROTEIN (ORF 2).
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 89384562.
RX GARRETT J.E., KNUTZON D.S., CARROLL D.;
RA "Composite transposable elements in the Xenopus laevis genome.";
RL MOL. CELL. BIOL. 9:3018-3027(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M26915; G214846; -
DR PIR: B32494; B32494.
DR PFAM: PF00078; rvt. 1.
KW HYPOTHETICAL PROTEIN; TRANSPOSABLE ELEMENT.
SQ SEQUENCE 1308 AA; 149577 MW; BE583757 CRC32;
Query Match 54.6%; Score 59; DB 1; Length 1308;
Best Local Similarity 58.3%; Pred. No. 7.63e-01;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 194 VRDGHVSQSRID 205
QY 25 VKGHVGQDRV 36
RESULT 4
ID VPH_BPHP1 STANDARD; PRT; 925 AA.
AC P51735;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PROBABLE TAIL FIBRE PROTEIN (ORF31).
OS BACTERIOPHAGE HP1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=HP1C1;
RX MEDLINE: 96279738.
RA ESPOSITO D., FITZMAURICE W.P., BENJAMIN R.C., GOODMAN S.D.,
RA WALDMAN A.S., SCOCCA J.J.;
RT "The complete nucleotide sequence of bacteriophage HP1 DNA.";
RL NUCLEIC ACIDS RES. 24:2360-2368(1996).
CC -1- SIMILARITY: SOME, TO PHAGE P22 PROTEIN H.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U24159; G1046260; -.

KW FIBER PROTEIN.

SO SEQUENCE 925 AA; 102001 MW; 538040CC CRC32;

Query Match 53.7%; Score 58; DB 1; Length 925;

Best Local Similarity 38.5%; Pred. No. 1.22e+00;

Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 813 VMGNVSSGRINI 825

QY 25 VMKGVGQDRVDF 37

RESULT 5

ID IF34 YEAST STANDARD; PRT; 347 AA.

AC P40217;

DT 01-FEB-1995 (REL. 31, CREATED)

DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 DELTA SUBUNIT (EIF-3 DELTA)

DE (EIF3 P39) (TRANSLATION INITIATION FACTOR EIF3, P39 SUBUNIT).

GN TIF34 OR YMR146C OR YMR375.16C.

OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;

OC SACCHAROMYCETACEAE; SACCHAROMYCES.

[1]

RP SEQUENCE FROM N.A.

RC STRAIN-W303D.

RX MEDLINE; 97127370.

RA NARANDA T., KAINUMA M., MACMILLAN S.E., HERSHEY J.W.B.;

RT "The 39-kilodalton subunit of eukaryotic translation initiation

factor 3 is essential for the complex's integrity and for cell

viability in *Saccharomyces cerevisiae*."

RL MOL. CELL. BIOL. 17:1145-1153(1997).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C / AB972;

RA BADCOCK K., CHURCHER C., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;

RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.

CC -!- FUNCTION: EIF-3 DISOCIATES RIBOSOMES, PROMOTES INITIATOR MET-TRNA

AND MRNA BINDING.

CC -!- SUBUNIT: EIF-3 IS COMPOSED OF UP TO 8 DIFFERENT SUBUNITS.

CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U56937; G1314862; -.

DR EMBL; 247071; G606445; -.

DR SGD; L0003329; TIF34.

DR PROSITE; PS00678; WD_REPEATS; 1.

DR PFAM; PF00400; G-beta; 2.

KW INITIATION FACTOR; PROTEIN BIOSYNTHESIS; REPEAT; WD REPEAT.

FT REPEAT 8 38

FT REPEAT 38 80

FT REPEAT 80 80

FT REPEAT 149 181

FT REPEAT 194 224

FT REPEAT 291 321

SO SEQUENCE 347 AA; 38755 MW; CF0D16F6 CRC32;

Query Match

Best Local Similarity 52.8%; Score 57; DB 1; Length 347;

Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Db 90 WKSPVPKRVSEF 101
QY 26 WKGVGQDRVDF 37

RESULT 6

ID RNPB RAT STANDARD; PRT; 419 AA.

AC P51607;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE RENIN-BINDING PROTEIN (RNPB).

GN RNPB.

OS RATUS NORVEGICUS (RAT).

OC EUKARYOTA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92138649.

RA INOUE H., TAKAHASHI S., FUKUI K., MIYAKE Y.;

RT "Genetic and molecular properties of human and rat renin-binding

proteins with reference to the function of the leucine zipper

motif."

RL J. BIOCHEM. 110:493-500(1991).

RN [2]

RP TISSUE SPECIFICITY.

RX MEDLINE; 93015776.

RA TADA M., TAKAHASHI S., MIYANO M., MIYAKE Y.;

RT "Tissue-specific regulation of renin-binding protein gene expression

in rats."

RL J. BIOCHEM. 112:175-182(1992).

CC -!- FUNCTION: PROTEIN THAT BINDS TO RENIN FORMING A PROTEIN COMPLEX

CALLED HIGH MOLECULAR WEIGHT (HMW) RENIN. THE COMPLEX IS

ENZYMATICALLY INACTIVE.

CC -!- SUBUNIT: HOMODIMER.

CC -!- TISSUE SPECIFICITY: KIDNEY, ADRENAL GLAND, BRAIN, LUNG, SPLEEN,

Ovary, Testis, and Heart.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D10233; D1001552; -.

DR DOMAIN 185 206 LEUCINE-ZIPPER.

SO SEQUENCE 419 AA; 48391 MW; ED41DC07 CRC32;

Query Match

Best Local Similarity 52.8%; Score 57; DB 1; Length 419;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 10 VMKORVGQE 18

QY 25 VMKGVGQD 33

RESULT 7

ID HFE_HUMAN STANDARD; PRT; 348 AA.

AC Q30201;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE HEREDITARY HAEMOCHROMATOSIS PROTEIN HLA-H PRECURSOR.

GN HFE OR HLAH.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A., VARIANT HH TYR-282, AND VARIANT ASP-63.
RX MEDLINE: 96331279.
RA FEDER J.N., GNIERKE A., THOMAS W., TSUCHIHASHI Z., RUDDY D.A.,
RA BASAVA A., DORMISHIAN F., DOMINGO R., ELLIS M.C. JR., FULLAN A.,
RA HINTON L.M., JONES N.L., KIMMEL B.E., KRONMAL G.S., LAUER P.,
RA LEE V.K., LOEB D.B., MAPA F.A., MCCLELLAND E., MEYER N.C.,
RA MINTIER G.A., MOELLER N., MOORE T., MORIKANG E., PRASS C.E.,
RA QUINTANA L., STARNES S.M., SCHATZMAN R.C., BRUNKE K.J.,
RA DRAYNA D.T., RISCH N.J., BACON B.R., WOLFF R.K.;
RT "A novel MHC class I-like gene is mutated in patients with hereditary
RT haemochromatosis.";
RT NAT. GENET. 13:399-409(1996).
[2]
RP SEQUENCE FROM N.A.
RA ALBIG W., BURMESTER N., BODE C., DOENECKE D., DRABENT B.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
[3]
RP SEQUENCE FROM N.A.
RA RUDDY D.A., KRONMAL G.S., LEE V.K., MINTIER G.A., QUINTANA L.,
RA DOMINGO R., MEYER N.C., IRRINKE A., MCCLELLAND E., FULLAN A.,
RA MAPA F.A., MOORE T., THOMAS W., LOEB D.B., HARMON C.,
RA TSUCHIHASHI Z., WOLFF R.K., SCHATZMAN R.C., FEDER J.N.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
[4]
RP SEQUENCE FROM N.A.
RA GASPARINI P.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
[5]
RP VARIANT HH TYR-282, AND VARIANT ASP-63.
RX MEDLINE: 97260408.
RA CARELLA M., D'AMBROSIO L., TOTARO A., GRIFA A., VALENTINO M.A.,
RA PIPERNO A., GIRELLI D., ROETTO A., FRANCO B., GASPARINI P.,
RA CAMASCHIELLA C.;
RT "Mutation analysis of the HLA-H gene in Italian hemochromatosis
RT patients.";
RL AM. J. HUM. GENET. 60:828-832(1997).
CC -!- FUNCTION: MAY BE A RECEPTOR FOR AN IRON-BINDING LIGAND.
CC -!- TISSUE SPECIFICITY: IN ALL TISSUES TESTED EXCEPT BRAIN.
CC -!- DISEASE: DEFECTS IN HFE ARE A CAUSE OF HEREDITARY HEMOCHROMATOSIS
CC (HH) HH IS AN AUTOSOMAL RECESSIVE INBORN DISORDER OF IRON
CC METABOLISM, FREQUENT AMONG CAUCASIANS. HH IS CHARACTERIZED BY
CC ABNORMAL INTESTINAL IRON ABSORPTION AND PROGRESSIVE INCREASE OF
CC TOTAL BODY IRON, WHICH RESULTS IN MIDLIFE IN CLINICAL
CC COMPLICATIONS INCLUDING CIRRHOSIS, CARDIOPATHY, DIABETES,
CC ENDOCRINE DYSFUNCTIONS, ARTHROPATHY, AND SUSCEPTIBILITY TO LIVER
CC CANCER. SINCE THE DISEASE COMPLICATIONS CAN BE EFFECTIVELY
CC PREVENTED BY REGULAR PHLEBOTOMIES, EARLY DIAGNOSIS IS MOST
CC IMPORTANT TO PROVIDE A NORMAL LIFE EXPECTANCY TO THE AFFECTED
CC SUBJECTS.
CC -!- SIMILARITY: TO MHC CLASS I ANTIGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U60319; G1469790; -
CC DR EMBL: 292910; E1287447; -
CC DR EMBL: U91328; G2088551; -
CC DR EMBL: Y09801; E291129; -
CC DR EMBL: Y09800; E291129; JOINED.
CC DR EMBL: Y09803; E291129; JOINED.
CC DR EMBL: Y09799; E291129; JOINED.
CC DR MIN: 235200; -
CC DR PROSITE: PS00290; IG_MHC; 1.
CC DR PFAM: PF00047; IG; 1.
CC DR PFAM: PF00129; MHC_I; 1.
CC DR HSP: P13760; 25EB.
KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL; DISEASE MUTATION;
KW POLYMORPHISM.

FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 348 HEREDITARY HAEMOCHROMATOSIS PROTEIN HLA-H.
FT DOMAIN 23 114 EXTRACELLULAR ALPHA-1.
FT DOMAIN 115 205 EXTRACELLULAR ALPHA-2.
FT DOMAIN 206 297 EXTRACELLULAR ALPHA-3.
FT TRANSMEM 298 306 CONNECTING PEPTIDE.
FT DOMAIN 307 330 POTENTIAL.
FT DOMAIN 331 348 CYTOPLASMIC TAIL.
FT DISULFID 124 187 BY SIMILARITY.
FT DISULFID 225 282 BY SIMILARITY.
FT CARBOHYD 110 110 POTENTIAL.
FT CARBOHYD 130 130 POTENTIAL.
FT CARBOHYD 234 234 POTENTIAL.
FT VARIANT 63 63 H -> D.
FT VARIANT 282 282 C -> Y (IN HH).
SQ SEQUENCE 348 AA; 40108 MW; D11DBF2 CRC32;

Query Match 51.9%; Score 56; DB 1; Length 348;
Best Local Similarity 53.8%; Pred. No. 3.07e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Db 136 WKYGDGQDHLEF 148
||| |||:::|
QY 26 WK-GHVGQDRVDF 37

RESULT 8
ID HFE_RAT STANDARD; PRT; 360 AA.
AC Q35799; Q35175;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HEREDITARY HAEMOCHROMATOSIS PROTEIN HLA-H HOMOLOG PRECURSOR (RT1-
DE CAFE).
GN HFE.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA BANASCH M.W., SCHAEFER H., SCHMIDT W.E.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
[2]
RN SEQUENCE OF 42-303 FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA SAWADA-HIRAI R., ROTHENBERG B.E.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
CC -!- FUNCTION: MAY BE A RECEPTOR FOR AN IRON-BINDING LIGAND.
CC -!- SIMILARITY: TO MHC CLASS I ANTIGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ001517; E340204; -
CC DR EMBL: AF008587; G2624957; -
CC DR PROSITE: PS00290; IG_MHC; 1.
CC DR PFAM: PF00047; IG; 1.
CC DR PFAM: PF00129; MHC_I; 1.
CC DR HSP: P30460; 1AGB.
KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 360 HEREDITARY HAEMOCHROMATOSIS PROTEIN HLA-H
FT HOMOLOG.
FT DOMAIN 26 127 EXTRACELLULAR ALPHA-1.
FT DOMAIN 128 218 EXTRACELLULAR ALPHA-2.
FT DOMAIN 219 310 EXTRACELLULAR ALPHA-3.

FT DOMAIN 311 319 CONNECTING PEPTIDE.
FT TRANSSEM 320 340 POTENTIAL.
FT DOMAIN 341 360 CYTOPLASMIC TAIL.
FT DISULFID 137 200 BY SIMILARITY.
FT DISULFID 238 295 BY SIMILARITY.
FT CARBOHYD 115 115 POTENTIAL.
FT CARBOHYD 143 143 POTENTIAL.
FT CARBOHYD 167 167 POTENTIAL.
FT CARBOHYD 247 247 POTENTIAL.
FT CONFLICT 198 198 R -> K (IN REF. 2).
SQ SEQUENCE 360 AA; 40988 MW; 43A04B0 CRC32;

Query Match 51.9%; Score 56; DB 1; Length 360;
Best Local Similarity 53.8%; Pred. No. 3.07e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Db 149 WKYGYDGDHLEF 161
||| |||::|

QY 26 WK-GHVGQDRVDF 37

RESULT 9 STANDARD; PRT; 472 AA.

ID EIBL_ADE41 PRT; 472 AA.

AC P10546;

DT 01-JUL-1989 (REL. 11, CREATED)

DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)

DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)

DE E1B PROTEIN, LARGE T-ANTIGEN (EARLY E1B 52 KD PROTEIN).

OS HUMAN ADENOVIRUS TYPE 41.

OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; ADENOVIRIDAE; MASTADENOVIRUS.

[1]

RP SEQUENCE FROM N.A.

RN MEDLINE; 88084437.

RA VAN LOON A.E.; LIGTENBERG M., REEMST A.M.C.B., SUSSENBACH J.S.,

RA ROZIJN T.H.;

RT "Structure and organization of the left-terminal DNA regions of

RT fastidious adenovirus types 40 and 41."

RL GENE 58:109-126(1987).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92230230.

RA ALLARD A., WADELL G.;

RT "The E1B transcription map of the enteric adenovirus type 41";

RL VIROLOGY 188:319-330(1992).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M18289; G209781; .

DR EMBL; M87544; G209892; .

DR PIR; F27333; WNA0F6.

KW EARLY PROTEIN

FT CONFLICT 54 55 MISSING (IN REF. 2).

FT CONFLICT 172 173 TV -> IL (IN REF. 2).

FT CONFLICT 234 234 S -> P (IN REF. 2).

SQ SEQUENCE 472 AA; 52157 MW; 7E9EE276 CRC32;

Query Match 51.9%; Score 56; DB 1; Length 472;
Best Local Similarity 66.7%; Pred. No. 3.07e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 270 WKGLVGONK 278
||| |||::|

QY 26 WKGHVGQDR 34

RESULT 10 STANDARD; PRT; 2255 AA.

ID RRPL_SV5 PRT; 2255 AA.

AC Q88434;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
(L PROTEIN).
GN L.
OS SIMIAN VIRUS 5 (STRAIN W3) (SV5).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92327825.
RA PARKS G.D., WARD C.D., LAMB R.A.;
RT "Molecular cloning of the NP and L genes of simian virus 5:
RT Identification of highly conserved domains in paramyxovirus NP and L
RT proteins.";
RL VIRUS RES. 22:259-279(1992).
CC -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P
CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M81721; G335113; .

DR EMBL; AF052755; G2981093; .

DR PFAM; PF00946; Paramyx_RNA_pol; 1.

KW TRANSFERASE; RNA-DIRECTED RNA POLYMERASE.

SQ SEQUENCE 2255 AA; 255923 MW; 223C1DAB CRC32;

Query Match 51.9%; Score 56; DB 1; Length 2255;
Best Local Similarity 50.0%; Pred. No. 3.07e+00;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 2027 WDGLKQERI 2036
|::|::|::|

QY 26 WKGHVGQDRV 35

RESULT 11

ID RRPL_SV5WR STANDARD; PRT; 2255 AA.

AC Q03396;

DT 01-FEB-1994 (REL. 28, CREATED)

DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)

(L PROTEIN).

GN L.

OS SIMIAN VIRUS 5 (STRAIN 21004-WR) (SV5).

OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;

OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92341049.

RA HIGUCHI Y., MIYAHARA Y., KAWANO M., TSURUDOME M., MATSUMURA H.,

RA KUSAGAWA S., KOMADA H., NISHIO M., ITO Y.;

RT "sequence analysis of the large (L) protein of simian virus 5.";

RL J. GEN. VIROL. 73:1005-1010(1992)

CC -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY

CC FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)

CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P

CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.

CC -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D13868; G222628; -.
DR PIR; J01532; J01532.
DR PFAM; PF00946; Paramyx RNA_pol. 1.
KW TRANSFERASE; RNA-DIRECTED RNA POLYMERASE.
SQ SEQUENCE 2255 AA; 25938 MW; 3A882A5F CRC32;

Query Match 51.9%; Score 56; DB 1; Length 2255;

Best Local Similarity 50.0%; Pred. No. 3.07e+00;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 2027 WGHKLQERI 2036

| | | | |

QY 26 WKGVGQDRV 35

RESULT 12
ID PDHX_MYXXA STANDARD; PRT; 270 AA.
AC P21159;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROBABLE PYRIDOXAMINE 5'-PHOSPHATE OXIDASE (EC 1.4.3.5) (PNP/PMP
OXIDASE) (FPA PROTEIN).
GN FPA.
OS MYXOCOCCUS XANTHUS.
OC BACTERIA; PROTEOBACTERIA; DELTA SUBDIVISION; MYXOBACTERIA;
OC MYXOCOCCALES; CYSTOBACTERINEAE; MYXOCOCCACEAE; MYXOCOCCUS.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 90094211.
RA HAGEN T.J., SHIMKETS L.J.;
RT "Nucleotide sequence and transcriptional products of the csf locus of
Myxococcus xanthus".
RL J. BACTERIOL. 172:15-23(1990).
[2]

RN FMN BINDING.
RX MEDLINE; 90094223.
RA SHIMKETS L.J.;
RT "The Myxococcus xanthus FprA protein causes increased flavin
biosynthesis in Escherichia coli."
RL J. BACTERIOL. 172:24-30(1990).
CC -!- FUNCTION: OXIDIZE PNP AND PMP INTO PYRIDOXAL 5'-PHOSPHATE (PLP)
(BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: PYRIDOXAMINE 5'-PHOSPHATE + H(2)O + O(2) =
PYRIDOXAL 5'-PHOSPHATE + NH(3) + H(2)O(2).
CC -!- COFACTOR: FMN.
CC -!- PATHWAY: DE NOVO SYNTHESIS OF PYRIDOXINE (VITAMIN B6) AND
PYRIDOXAL PHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE PYRIDOXAMINE 5'-PHOSPHATE OXIDASE
FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M29288; G150085; ALT_INIT.
DR PROSITE; PS01064; PYRIDOX_OXIDASE; 1.
DR PFAM; PF01243; Pyridox_oxidase; 1.
KW PYRIDOXINE BIOSYNTHESIS; OXIDOREDUCTASE; FLAVOPROTEIN; FMN.
SQ SEQUENCE 270 AA; 30465 MW; 4C546E53 CRC32;

Query Match 50.9%; Score 55; DB 1; Length 270;

Best Local Similarity 46.2%; Pred. No. 4.82e+00;
Matches 6; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Db 230 WSGFRVVPDRIEF 242

| | | | |

QY 26 WKGVGQDRVDF 37

RESULT 13
ID XYNZ_CLOTH STANDARD; PRT; 837 AA.
AC P10478;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ENDO-1,4-BETA-XYLANASE 2 PRECURSOR (EC 3.2.1.8) (XYLANASE 2)
DE (1,4-BETA-D-XYLAN XYLANOXYDROLASE 2).
GN XYNZ.
OS CLOSTRIDIUM THERMOCELLUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
[1]
RN SEQUENCE FROM N.A.
RX STRAIN=NCIB 10682;
RX MEDLINE; 89008072.
RA GREPINET O., CHEBROU M.-C., BEGUIN P.;
RT "Nucleotide sequence and deletion analysis of the xylanase gene
(xynZ) of Clostridium thermocellum".
RL J. BACTERIOL. 170:4582-4588(1988).
[2]
RN X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS) OF 515-837.
RX STRAIN=NCIB 10682;
RX MEDLINE; 95393242.
RA DOMINGUEZ R., SOUCHON H., SPINELLI S., DAUTER Z., WILSON K.S.,
RA CHAUVARX S., BEGUIN P., ALZARI P.M.;
RT "A common protein fold and similar active site in two distinct
families of beta-glycanases".
RL NAT. STRUCT. BIOL. 2:569-576(1995).
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
LINKAGES IN XYLANS.
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
CC -!- SIMILARITY: CONTAINS 1 XYNZ-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M22624; G144932; -.
DR PIR; A31842; A31842.
DR PDB; 1XYZ; 29-JAN-96.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 2.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PFAM; PF00331; glycosyl_hydro3; 1.
DR PFAM; PF00404; celCC; 2.
KW XLAN DEGRADATION; HYDROLASE; GLYCOSIDASE; REPEAT; SIGNAL;
3D-STRUCTURE.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 837 XYLANASE 2.
FT ACT_SITE 645 645 PROTON-DONOR.
FT ACT_SITE 754 754 NUCLEOPHILE.
FT DOMAIN 328 416 CELLULOSE-BINDING (POTENTIAL).
FT DOMAIN 430 487 2 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 430 453 1.
FT REPEAT 464 487 2.
FT DISULFID 783 789 BY SIMILARITY.

```
SQ SEQUENCE 837 AA: 92262 MW; D1B5D725 CRC32;
Query Match 50.9%; Score 55; DB 1: Length 837;
Best Local Similarity 38.5%; Pred.No. 4.82e+00;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 658 IWRNVIGQDYLDY 670
:|: :||| :|:
QY 25 VKGHVGDQDVDF 37

RESULT 14
ID SPCO_MOUSE STANDARD; PRT: 2363 AA.
AC Q62261;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE SPECTRIN BETA CHAIN, BRAIN (SPTBN1, NON-ERYTHROID BETA CHAIN)
DE (FODRIN BETA CHAIN) (SPTBN1).
GN SPTB2 OR SPNB2 OR SPNB-2.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C.
RX MEDLINE: 93240985.
RA MA Y., ZIMMER W.E., RIEDERER B.M., GOODMAN S.R.;
RT "The complete amino acid sequence for brain beta spectrin (beta fodrin): relationship to globin sequences.";
RL BRAIN RES. MOL. BRAIN RES. 18:87-89(1993).
RN [2]
RP STRUCTURE BY NMR OF 2199-2304.
RX MEDLINE: 94268558.
RA MACIAS M.J., MUSACCHIO A., PONSTINGL H., NILGES M., SARASTE M.,
RA OSCHKINAT H.;
RT "Structure of the pleckstrin homology domain from beta-spectrin.";
RL NATURE 369:675-677(1994).
RN [3]
RP STRUCTURE BY NMR OF 2199-2304.
RX MEDLINE: 97342712.
RA NILGES M., MACIAS M.J., O'DONOGHUE S.I., OSCHKINAT H.;
RT "Automated NORSY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin.";
RL J. MOL. BIOL. 269:408-422(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 2199-2304.
RX MEDLINE: 96030773.
RA HYVOENEN M., MACIAS M.J., NILGES M., OSCHKINAT H., SARASTE M.,
RA WILMANN M.;
RT "Structure of the binding site for inositol phosphates in a PH domain.";
RL EMBO J. 14:4676-4681(1995).
CC -!- FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION, INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS THUS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE CYTOSKELETON AT THE MEMBRANE.
CC -!- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO TETRAMERS.
CC -!- DOMAIN: SPECTRIN-LIKE PROTEINS HAVE FIVE DOMAINS: (1) N-TERMINAL DOMAIN (N), (2) DOMAIN BETWEEN THE N-TERMINAL AND MIDDLE DOMAIN (NM), (3) MIDDLE DOMAIN (M), (4) DOMAIN BETWEEN THE MIDDLE AND C-TERMINAL DOMAIN (MC), (5) C-TERMINAL DOMAIN (C).
CC NM AND MC DOMAINS ARE COMPOSED OF TYPICAL SPECTRIN 106 RESIDUES REPEATS (1-8 FOR NM AND 12-19 FOR MC) AND ARE HOMOLOGOUS TO EACH OTHER. N, M, AND C DOMAINS ARE COMPOSED OF SEQUENCES THAT DO NOT FORM TYPICAL SPECTRIN REPEATS.
CC -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL: M74773; G409226;
PDB: 1BTN; 08-MAR-96.
PDB: 1MPH; 16-JUN-97.
DR MGD; MGI:98388; SPNB2.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PFAM; PF00169; PH; 1.
DR PFAM; PF00307; actinin-binding; 1.
DR PFAM; PF00435; spectrin; 18.
KW CYTOSKELETON; MEMBRANE; REPEAT; ACTIN-BINDING; CAPPING PROTEIN;
KW 3D-STRUCTURE.
DOMAIN 1 272 ACTIN-BINDING (BY SIMILARITY).
FT REPEAT 273 384 1.
FT REPEAT 385 498 2.
FT REPEAT 499 608 3.
FT REPEAT 609 714 4.
FT REPEAT 715 819 5.
FT REPEAT 820 925 6.
FT REPEAT 926 1032 7.
FT REPEAT 1033 1139 8.
FT REPEAT 1140 1245 9.
FT REPEAT 1246 1350 10.
FT REPEAT 1351 1462 11.
FT REPEAT 1463 1562 12.
FT REPEAT 1563 1668 13.
FT REPEAT 1669 1775 14.
FT REPEAT 1776 1881 15.
FT REPEAT 1882 1987 16.
FT REPEAT 1988 2132 17.
FT DOMAIN 2196 2306 PH.
SQ SEQUENCE 2363 AA; 274420 MW; 3E31781D CRC32;
Query Match 50.9%; Score 55; DB 1: Length 2363;
Best Local Similarity 56.7%; Pred.No. 4.82e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1748 GNIGQERVD 1756
:|: :||| :|:
QY 28 GHVGDQDVDF 36

RESULT 15
ID SPCO_HUMAN STANDARD; PRT: 2364 AA.
AC Q01082; Q16057;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE SPECTRIN BETA CHAIN, BRAIN (SPTBN1, NON-ERYTHROID BETA CHAIN)
DE (FODRIN BETA CHAIN) (SPTBN1).
GN SPTB2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE: 92406787.
RA HU R.J., WATANABE M., BENNETT V.;
RT "Characterization of human brain cDNA encoding the general isoform of beta-spectrin.";
RL J. BIOL. CHEM. 267:18715-18722(1992).
RN [2]
RP SEQUENCE OF 293-1544 FROM N.A.
```

RX MEDLINE; 94010920.
RA CHANG J.G., SCARPA A., EDDY R.L., BYERS M.G., HARRIS A.S.,
RA MORROW J.S., WATKINS P., SHOWS T.B., FORGET B.G.;
RT "Cloning of a portion of the chromosomal gene and cDNA for human
RT beta-fodrin, the nonerythroid form of beta-spectrin.";
RL GENOMICS 17:287-293(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 173-280.
RX MEDLINE; 97307247.
RA CARUGO K.D., BANELOS S., SARASTE M.;
RT "Crystal structure of a calponin homology domain.";
RL NAT. STRUCT. BIOL. 4:175-179(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 173-281.
RA BANELOS S., SARASTE M., CARUGO K.D.;
RL SUBMITTED (JUL-1998) TO THE PDB DATA BANK.
CC -!- FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,
CC INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS
CC THIS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE
CC CYTOSKELETON AT THE MEMBRANE.
CC -!- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
CC ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
CC TETRAMERS.
CC -!- DOMAIN: SPECTRIN-LIKE PROTEINS HAVE FIVE DOMAINS:
CC (1) N-TERMINAL DOMAIN (N),
CC (2) DOMAIN BETWEEN THE N-TERMINAL AND MIDDLE DOMAIN (NM),
CC (3) MIDDLE DOMAIN (M),
CC (4) DOMAIN BETWEEN THE MIDDLE AND C-TERMINAL DOMAIN (MC),
CC (5) C-TERMINAL DOMAIN (C).
CC NM AND MC DOMAINS ARE COMPOSED OF TYPICAL SPECTRIN 106 RESIDUES
CC REPEATS (1-8 FOR NM AND 12-19 FOR MC) AND ARE HOMOLOGOUS TO EACH
CC OTHER. N, M, AND C DOMAINS ARE COMPOSED OF SEQUENCES THAT DO NOT
CC FORM TYPICAL SPECTRIN REPEATS.
CC -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M96803; G338443; -
DR EMBL; S65762; G425553; -
DR PDB; 1AA2; 04-FEB-98.
DR PDB; 1BKR; 15-JUL-98.
DR MIM; 182790; -
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PFAM; PF00169; PH; 1.
DR PFAM; PF00307; actinin-binding; 1.
DR PFAM; PF00435; spectrin; 17.
KW CYTOSKELETON; MEMBRANE; REPEAT; ACTIN-BINDING; CAPPING PROTEIN;
3D-STRUCTURE.
FT DOMAIN 1 272 ACTIN-BINDING (BY SIMILARITY).
FT REPEAT 273 384 1.
FT REPEAT 385 498 2.
FT REPEAT 499 608 3.
FT REPEAT 609 714 4.
FT REPEAT 715 819 5.
FT REPEAT 820 925 6.
FT REPEAT 926 1032 7.
FT REPEAT 1033 1139 8.
FT REPEAT 1140 1245 9.
FT REPEAT 1246 1350 10.
FT REPEAT 1351 1462 11.
FT REPEAT 1463 1562 12.
FT REPEAT 1563 1668 13.
FT REPEAT 1669 1775 14.
FT REPEAT 1776 1881 15.

FT REPEAT 1882 1987 16.
FT REPEAT 1988 2133 17.
FT DOMAIN 2197 2307 PH.
FT CONFLICT 1411 1411 H -> D (IN REF. 2).
SQ SEQUENCE 2364 AA; 274629 MW; D45B9267 CRC32;
Query Match 50.9%; Score 55; DB 1; Length 2364;
Best Local Similarity 66.7%; Pred. No. 4.82e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 1748 GNIGQERVD 1756
Qy 28 GHVGQDRVD 36
|::|::|::|
|::|::|::|
Search completed: Thu Jul 8 18:08:11 1999
Job time : 11 secs.

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:08:29 1999; Maspar time 5.42 Seconds
Tabular output not generated. 130.942 Million cell updates/sec

Title: >US-09-041-236-2
Description: (25-37) from US09041236.pep (3 of 45)
Perfect Score: 108
Sequence: 1 VWKGHVQDQDRVDF 13

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 26.578; Variance 32.249; scale 0.824

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	108	100.0	666	4	075326 SEMAPHORIN L.	7.63e-13
2	65	60.2	403	11	P97306 STAC.	5.70e-02
3	63	58.3	402	4	Q99469 STAC.	1.57e-01
4	58	53.7	234	5	Q22139 T04A8.10 PROTEIN.	1.81e+00
5	57	52.8	250	2	O06227 YF1H.	2.91e+00
6	57	52.8	254	2	P73626 HYPOTHETICAL 28.5 KD P	2.91e+00
7	57	52.8	353	1	O28918 HYPOTHETICAL 39.3 KD P	2.91e+00
8	57	52.8	595	3	P78704 ACR-2 PROTEIN.	2.91e+00
9	56	51.9	246	4	O75931 HEMOCHROMATOSIS SPLICE	4.64e+00
10	56	51.9	260	4	O75930 HEMOCHROMATOSIS SPLICE	4.64e+00
11	56	51.9	334	4	O75929 HEMOCHROMATOSIS SPLICE	4.64e+00
12	56	51.9	1321	4	O75129 KIAA0634 PROTEIN (FRAG	4.64e+00
13	55	50.9	312	5	O17812 F15A4.9 PROTEIN.	7.36e+00
14	55	50.9	313	2	P94522 ARABINAN-ENDO 1,5-ALPH	7.36e+00
15	55	50.9	322	10	Q41419 SUCROLYTIC ENZYME/FERR	7.36e+00
16	55	50.9	324	2	O07078 ENDO-ARABINASE.	7.36e+00
17	55	50.9	475	3	O59812 GLUCOSE-6-PHOSPHATE DE	7.36e+00
18	55	50.9	707	6	Q28297 BETA-SPECTRIN (FRAGMEN	7.36e+00
19	55	50.9	1224	2	P95629 PUTA GENE.	7.36e+00
20	55	50.9	1228	2	Q44334 PROLINE DEHYDROGENASE.	7.36e+00

21	55	50.9	3623	11	070244	INTRINSIC FACTOR-B12 R	7.36e+00
22	54	50.0	293	2	067668	THIOSULFATE SULFURTRAN	1.16e+01
23	54	50.0	317	9	080052	ORF 13	1.16e+01
24	54	50.0	469	2	P73738	HYPOTHETICAL 52.5 KD P	1.16e+01
25	54	50.0	499	2	051842	OUTER MEMBRANE PROTEIN	1.16e+01
26	54	50.0	577	4	060907	TRANSUDIN (BETA) LIKE	1.16e+01
27	54	50.0	1220	14	041894	POL POLYPROTEIN.	1.16e+01
28	53	49.1	230	2	084183	ABC TRANSPORT ATPASE (1.82e+01
29	53	49.1	300	2	053925	PUTATIVE TRANSMEMBRANE	1.82e+01
30	53	49.1	607	10	P93622	POLIPHENOL OXIDASE (EC	1.82e+01
31	53	49.1	760	10	081732	HYPOTHETICAL 85.3 KD P	1.82e+01
32	52	48.1	199	2	034723	SENSOR REGULATOR.	2.82e+01
33	52	48.1	402	6	Q95331	N-ACYL-D-GLUCOSAMINE 2	2.82e+01
34	52	48.1	415	2	084824	TYROSINE TRANSPORT.	2.82e+01
35	52	48.1	460	10	P93018	TYROSYL-TRNA SYNTHETAS	2.82e+01
36	52	48.1	669	9	080185	PUTATIVE MINOR STRUCTU	2.82e+01
37	52	48.1	669	2	034077	ORF46.	2.82e+01
38	52	48.1	838	6	019056	FERTILIN ALPHA-1 (FRAG	2.82e+01
39	52	48.1	943	1	030320	ATP-DEPENDENT RNA HELI	2.82e+01
40	52	48.1	1006	3	074458	PUTATIVE TRANSCRIPTION	2.82e+01
41	52	48.1	1732	2	007442	LYSINE-SPECIFIC CYSTEI	2.82e+01
42	52	48.1	1732	2	Q52050	LYSINE-SPECIFIC CYSTEI	2.82e+01
43	52	48.1	1732	2	Q51817	PORPHYRAIN.	2.82e+01
44	52	48.1	1756	10	080975	F26C24.9 PROTEIN.	2.82e+01
45	52	48.1	3085	14	070710	POLYPROTEIN.	2.82e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	666	AA.
ID	075326				
AC	075326;				
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L.				
GN	SENA.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;				
OC	CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98389619.				
RA	LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;				
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA				
RT	viruses."				
RL	GENOMICS 51:340-350(1998).				
DR	EMBL; AF030698; G3523115; "				
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;				

Query Match 100.0%; Score 108; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 7.63e-13;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 57 VWKGHVQDQDRVDF 69

Qy 25 VWKGHVQDQDRVDF 37

RESULT 2
ID P97306
AC P97306;

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DE 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE STAC.
GN STAC.

OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H; TISSUE=BRAIN;

RX MEDLINE: 97115677.
RA SUZUKI H., KAWAI J., TAGA C., YAOI T., HARA A., HIROSE K.,
RA HAYASHIZAKI Y., WATANABE S.;
RT "Stac, a novel neuron-specific protein with cysteine-rich and SH3
domains.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 229:902-909(1996).
DR EMBL: D86639; D1013839; -.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00130; DAG-PE-bind; 1.
SQ SEQUENCE 403 AA; 44319 MW; B152DCB7 CRC32;

Query Match 60.2%; Score 65; DB 11; Length 403;
Best Local Similarity 66.7%; Pred. No. 5.70e-02;
Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Db 324 WKGKI-QDRVGF 334
QY 26 WKGHVQDQDVDF 37

RESULT 3
ID Q99469 PRELIMINARY; PRT; 402 AA.
AC Q99469;
DT 01-MAY-1997 (TREMREL. 03, CREATED)
DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE STAC.
GN STAC.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97115677.
RA SUZUKI H., KAWAI J., TAGA C., YAOI T., HARA A., HIROSE K.,
RA HAYASHIZAKI Y., WATANABE S.;
RT "Stac, a novel neuron-specific protein with cysteine-rich and SH3
domains.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 229:902-909(1996).
DR EMBL: D86640; D1013840; -.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00130; DAG-PE-bind; 1.
SQ SEQUENCE 402 AA; 44553 MW; 6FBB4B41 CRC32;

Query Match 58.3%; Score 63; DB 4; Length 402;
Best Local Similarity 58.3%; Pred. No. 1.57e-01;
Matches 7; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Db 323 WKGKI-QDRIGF 333
QY 26 WKGHVQDQDVDF 37

RESULT 4
ID Q22139 PRELIMINARY; PRT; 234 AA.
AC Q22139;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE T0A8-10 PROTEIN.
GN T0A8-10.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA PALMER S.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; 235663; E1348906; -.
SQ SEQUENCE 234 AA; 27015 MW; F4D69E2C CRC32;

Query Match 53.7%; Score 58; DB 5; Length 234;
Best Local Similarity 33.3%; Pred. No. 1.81e+00;
Matches 4; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 144 WKEEGEEQVEL 155
QY 26 WKGHVQDQDVDF 37

RESULT 5
ID O06227 PRELIMINARY; PRT; 250 AA.
AC O06227;
DT 01-JUL-1997 (TREMREL. 04, CREATED)
DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE YFII.
GN YFII.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORINEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA BADCOCK K., CHURCHER C.M.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA BARRELL B.G., RAJANDREAM M.A., PARKHILL J.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE: 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus.
Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
leprae.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
DR EMBL; 295388; E315955; -.
SQ SEQUENCE 250 AA; 25964 MW; B77ED924 CRC32;

Query Match 52.8%; Score 57; DB 2; Length 250;
Best Local Similarity 46.2%; Pred. No. 2.91e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 71 VMNVQVHGDVREL 83
QY 25 WKGHVQDQDVDF 37

RESULT 6
ID P73626 PRELIMINARY; PRT; 254 AA.
AC P73626;
DT 01-FEB-1997 (TREMREL. 02, CREATED)
DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)

DT	01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE	HYPOTHETICAL 28.5 KD PROTEIN.
OS	SYNECHOCYSTIS SP. (STRAIN FCC 6803).
OC	BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RP	[1]
RP	SEQUENCE FROM N.A.
RP	STRAIN=PCC6803;
RA	TABATA S.;
RL	[2]
RP	SEQUENCE FROM N.A.
RP	STRAIN=PCC6803;
RC	MEDLINE: 97061201.
RA	KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA	MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA	HOSHOUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA	SHIMO S., TAKEUCHI C., WADA T.; WATANABE A., YAMADA M., YASUDA M.,
RA	TABATA S.;
RT	"Sequence analysis of the genome of the unicellular cyanobacterium
RT	Synechocystis sp. PCC6803. II. Sequence determination of the entire
RT	genome and assignment of potential protein-coding regions.";
RL	DNA RES. 3:109-136(1996).
DR	EMBL: D90908; D1018404; -
DR	PFAM: PF00481; PP2C; 2.
SW	HYPOTHETICAL PROTEIN.
SK	SEQUENCE 254 AA; 28472 MW; 8524DECC CRC32;

```
Query Match      52.8%; Score 57; DB 2; Length 254;
Best Local Similarity 50.0%; Pred. No. 2.91e+00;
Matches      5; Conservative      4; Mismatches 1; Indels 0; Gaps 0;
```

```

Db      119 WRAHVGDSRL 128
      1::111:1:
Oy      26 WKGHVGDPRV 35

```

[illegible]

Query Match 52.8%; Score 57; DB 1; Length 353;
Best Local Similarity 55.6%; Pred. No. 2.91e+00;
Matches 5; Conservative 3; Mismatches 1; Indels

```

Db      119 WNGEVGEER 127
      I:| ||::|
Qy      26 WKGHVGQDR 34

RESULT      8
ID      P78704      PRELIMINARY;      PRT;      595 AA.
AC      P78704;
DT      01-MAY-1997 (TREMBLREL. 03, CREATED)
DT      01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT      01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE      ACR-2 PROTEIN.
GN      ACR-2.
OS      NEUROSPORA CRASSA.
OC      EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC      SORDARIALES; SORDARIACEAE; NEUROSPORA.
RN      [1]
RC      SEQUENCE FROM N.A.
RC      STRAIN=ACR-2 (KH2);
RX      MEDLINE: 96283814.
RA      AKIYAMA M., NAKASHIMA H.;
RT      "Molecular cloning of the acr-2 gene which controls acriflavine
RT      sensitivity in Neurospora crassa.";
RL      BIOCHIM. BIOPHYS. ACTA 1307:187-192(1996).
CC      -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC      -1- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR (C
CC      DOMAIN.
DR      EMBL: D45893; D1008903; -.
DR      PROSITE: PS00463; ZN2_CYS6_FUNGAL_1; 1.
DR      PFAM: PF001172; Zn_c1us; 1.
KW      TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN; ZINC;
KW      METAL-BINDING.
SQ      SEQUENCE 595 AA; 65291 MW; 3E789005 CRC32;

```

Query Match	52.8%	Score 57;	DB 3;	Length 595;
Best Local Similarity	66.7%;	Pred. No.	2.91e+00;	
Matches	6;	Conservative	1;	Mismatches 2;
		Indels	0;	Gaps 0;

```

Db      314 WKAHVGAAR 322
      ||:| | | |
Qv      26 WKGHVGDGR 34

```

RESULT	9		
ID	O75931	PRELIMINARY;	PRT; 246 AA.
AC	O75931;		
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	HEMOCROMATOSIS SPLICER VARIANT DELE214E4.		
GN	HFE.		
OS	HOMO SAPIENS (HUMAN).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;		
OC	CATARRHINI; HOMINIDAE; HOMO.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	RHODES D.A., TROWSDALE J.;		
RT	"Alternate splice variants of the haemochromatosis gene Hfe.";		
RL	SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.		
DR	EMBL; AF079409; G369511; -		
SQ	SEQUENCE 246 AA; 28150 MW; 99034D2C CRC32;		

Query Match	51.9%	Score 56;	DB 4;	Length 246;
Best Local Similarity	53.8%;	Pred. No. 4	64e+00;	
Matches	7;	Conservative	3;	Mismatches 2;
		Indels	1;	Gaps 1;

D _b	48	WKYGYDGDHLEF	60
		::	
Q _v	26	WK-GHVGODRVDF	37

RESULT 10

```
ID O75930 PRELIMINARY: PRT: 260 AA.
AC O75930;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE HEMOCHROMATOSIS SPLICE VARIANT DELE2.
GN HFE.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA RHODES D.A., TROWSDALE J.;
RT "Alternate splice variants of the haemochromatosis gene hfe.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF079408; G3695109; -.
SQ SEQUENCE 260 AA; 29633 MW; 1AA649F CRC32;

Query Match 51.9%; Score 56; DB 4; Length 260;
Best Local Similarity 53.8%; Pred. No. 4.64e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Db 48 WKYGDGQDHLF 60
QY 26 WK-GHVGQDRVDF 37
||| |:::|

RESULT 11
ID O75929 PRELIMINARY: PRT: 334 AA.
AC O75929;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE HEMOCHROMATOSIS SPLICE VARIANT DEL14E4.
GN HFE.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA RHODES D.A., TROWSDALE J.;
RT "Alternate splice variants of the haemochromatosis gene hfe.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF079407; G3695107; -.
SQ SEQUENCE 334 AA; 38625 MW; CABC425E CRC32;

Query Match 51.9%; Score 56; DB 4; Length 334;
Best Local Similarity 53.8%; Pred. No. 4.64e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Db 136 WKYGDGQDHLF 148
QY 26 WK-GHVGQDRVDF 37
||| |:::|

RESULT 12
ID O75129 PRELIMINARY: PRT: 1321 AA.
AC O75129;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE KIAA0634 PROTEIN (FRAGMENT).
GN KIAA0634.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-BRAIN;
RX MEDLINE; 98403880.
RA ISHIKAWA K., NAGASE T., SUYAMA M., MIYAJIMA N., TANAKA A., KOTANI H.,
RA NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
```

```
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA RES. 5:169-176(1998).
DR EMBL; AB014534; D1032570; -.
FT NON_TER 1
SQ SEQUENCE 1321 AA; 145424 MW; 4B1721D3 CRC32;

Query Match 51.9%; Score 56; DB 5; Length 1321;
Best Local Similarity 60.0%; Pred. No. 4.64e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 359 GHPGEEKVDF 368
QY 28 GHVGQDRVDF 37
||| |:::|

RESULT 13
ID O17812 PRELIMINARY: PRT: 312 AA.
AC O17812;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE F15A4.9 PROTEIN.
DE F15A4.9
GN F15A4.9.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA SIMS M.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; 281062; E1345513; -.
SQ SEQUENCE 312 AA; 35599 MW; 217BEF77 CRC32;

Query Match 50.9%; Score 55; DB 5; Length 312;
Best Local Similarity 25.0%; Pred. No. 7.36e+00;
Matches 3; Conservative 8; Mismatches 1; Indels 0; Gaps 0;

Db 72 LWKAEMSENRIE 83
QY 25 VMKGHVQDRVD 36
||| |:::|

RESULT 14
ID P94522 PRELIMINARY: PRT: 313 AA.
AC P94522;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ARABINAN-ENDO 1,5-ALPHA-L-ARABINASE.
GN ABNA.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA GOETHEL S.F., SCHMID R., WIPAT A., CARTER N.M., EMMERSON P.T.,
```

RA HARWOOD C.R., MARAHIEL M.A.;
RL EUR. J. BIOCHEM. 244:59-65(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE: 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOFIN A., BORCHERT S.,
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA ENTIAN K.D., ERRINGTON J., FABBET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUNA S., GALIZZI A., GALLERON N.,
RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISEPPI G., GUY B.J., HAGA K., HALECH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA JORIS B., KARAWATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA KOBAYASHI Y., KOETTER P., KONIGSTEIN G., KROGH S., KUMANO M.,
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA NOONE D., O'REILLY M., OGAWA K., OGAWARA A., OUDEGA B., PARK S.H.,
RA PARRO V., POHL T.M., PORTELELLA D., PORWOLLIK S., PRESCOTT A.M.,
RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SEKIGUCHI J., SEKOWSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDI B.,
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUNSTEIN E., YOSHIKAWA H., DANCHIN A.,
RT "The complete genome sequence of the gram-positive bacterium *Bacillus*
RT *subtilis*;"
RL NATURE 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: 275208; E1165305;
DR EMBL: 299118; E1184130;
SQ SEQUENCE 313 AA; 34281 MW; 7FCB3003 CRC32;

Query Match 50.9%; Score 55; DB 2; Length 313;
Best Local Similarity 63.6%; Pred. No. 7.36e+00;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 280 WKGGGQDIVN 290
QY 26 WKGHVGQDRVD 36
||| ||| |
||| ||| |

RESULT 15
ID Q41419 PRELIMINARY; PRT; 322 AA.
AC Q41419;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE SUCROLYTIC ENZYME/FERREDOXIN HOMOLOG.
OS SOLANUM TUBEROSUM (POTATO).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANALES; SOLANACEAE; SOLANUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95046297.
RA MACHRAY G.C., BURCH L., HEDLEY P.E., DAVIES H.V., WAUGH R.;
RT "Characterisation of a complementary DNA encoding a novel plant
RT enzyme with sucrolytic activity;"
RL FEBS LETT. 354:123-127(1994).

DR EMBL: S74161; G707001;
DR MENDEL: 17478; SOLtu:2917; mml7478.
SQ SEQUENCE 322 AA; 35552 MW; 3088B9F2 CRC32;

Query Match 50.9%; Score 55; DB 10; Length 322;
Best Local Similarity 41.7%; Pred. No. 7.36e+00;
Matches 5; Conservative 6; Mismatches 0; Indels 1; Gaps 1;

Db 214 LWRGQMGQYEV 225
QY 25 VMKGHVQD-DRV 35
:|:|:|:|:|:|

Search completed: Thu Jul 8 18:08:49 1999
Job time : 20 secs.

THIS PAGE BLANK (USPTO)

WQSRFH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:11:45 1999; MasPar time 11.14 Seconds
34.358 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
(42-59) from US09041236.pep (4 of 45)
Description: 136
Perfect Score: 1
Sequence: 1 PHTVLFEPGSSSVWVG 18

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 20.558; Variance 70.547; scale 0.291

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	63	46.3	3722	R10145	Cephalosporin antibiotic	4.19e+01
2	62	45.6	165	R42838	Prochymosin N-termina	5.20e+01
3	62	45.6	276	P80846	Sequence of proteinas	5.20e+01
4	62	45.6	365	P30603	Sequence encoded by p	5.20e+01
5	62	45.6	365	R05080	Sequence of calf pro-	5.20e+01
6	62	45.6	365	P94144	Prochymosin.	5.20e+01
7	62	45.6	375	P40078	Sequence encoded by p	5.20e+01
8	62	45.6	379	P30013	Sequence encoded by v	5.20e+01
9	62	45.6	380	P20038	pre-prorennin-A prote	5.20e+01
10	62	45.6	381	P40559	Sequence of a polypep	5.20e+01
11	62	45.6	381	P30086	Sequence encoded by c	5.20e+01
12	62	45.6	381	P30446	Sequence encoded by p	5.20e+01
13	62	45.6	381	R20730	Prochymosin (prorenni	5.20e+01
14	62	45.6	381	P40218	Sequence of rennin en	5.20e+01
15	62	45.6	450	P94370	Sequence encoded by B	5.20e+01
16	62	45.6	458	P94376	BamHI/SalI insert of	5.20e+01

17	61	44.9	24	9	R47014	Cathepsin E position	6.45e+01
18	61	44.9	406	3	P50135	Sequence of pre-pro-r	6.45e+01
19	61	44.9	406	26	W23244	Human renin.	6.45e+01
20	61	44.9	563	16	R86989	Alpha-1,3/4-fucosidas	6.45e+01
21	60	44.1	412	36	W71369	Death associated prot	7.99e+01
22	60	44.1	412	14	R74207	Human death associate	7.99e+01
23	60	44.1	585	2	P70143	BAR1 gene product.	7.99e+01
24	60	44.1	587	3	R20109	BAR1 barrier protease	7.99e+01
25	60	44.1	587	3	R13383	Barrier protein.	7.99e+01
26	60	44.1	596	32	W48785	Thyroid peroxidase de	7.99e+01
27	60	44.1	626	32	W48784	Thyroid peroxidase de	7.99e+01
28	60	44.1	712	2	R12124	Bovine lactoperoxidas	7.99e+01
29	60	44.1	784	32	W48783	Thyroid peroxidase de	7.99e+01
30	60	44.1	852	32	W48782	Thyroid peroxidase de	7.99e+01
31	60	44.1	881	32	W48791	Thyroid peroxidase de	7.99e+01
32	60	44.1	924	6	R35445	Human TPO lacking ami	7.99e+01
33	60	44.1	933	8	R44615	Human thyroid peroxid	7.99e+01
34	60	44.1	933	6	R32875	Human TPO.	7.99e+01
35	60	44.1	933	13	R75689	Human thyroid peroxid	7.99e+01
36	60	44.1	948	32	W48781	Thyroid peroxidase.	7.99e+01
37	58	42.6	323	18	R99487	PcpD.	1.22e+02
38	58	42.6	504	28	W35008	Vibrio harveyi endogl	1.22e+02
39	57	41.9	1464	38	W85576	Human N-methyl-D-aspa	1.51e+02
40	57	41.9	1464	18	R80970	Human excitatory amin	1.51e+02
41	57	41.9	1464	8	R42054	Glutamic acid recepto	1.51e+02
42	57	41.9	1464	10	R55329	Human NMDA R2A recept	1.51e+02
43	57	41.9	1464	13	R44192	Rat NMDA receptor sub	1.51e+02
44	57	41.9	1464	12	R66039	Human N-methyl-D-aspa	1.51e+02
45	56	41.2	356	17	R88210	Human serum paraoxona	1.86e+02

ALIGNMENTS

RESULT 1
ID R10145 standard; Protein; 3722 AA.

AC R10145;
DT 27-MAR-1991 (first entry)
DE Cephalosporin antibiotic biosynthetic enzyme #1.
KW cephalosporin; antibiotic;
KW S-(L-alpha-aminoadipyl)-L-cysteinyl-D-; valine synthetase;
KW isopenicillin N synthetase; isopenicillin N epimerase;
KW deacetoxycephalosporin C synthetase; beta-lactamase;
KW deacetoxycephalosporin C hydroxylase.
OS Lyso bacter lactamgenus.
PN J02291274-A.
PD 03-DEC-1990.
PF 10-JAN-1990; 003762.
PR 01-FEB-1989; JP-024710.
PR 10-JAN-1990; JP-003762.
PA (TAKE) TAKEDA CHEMICAL IND KK.
DR WPI; 91-018854/03.
DR N-PSDB; Q10190.
PT Prepn. of cephalosporin series antibiotics - comprises culturing
PT transformant of microbe transformed by plasmid contg. new DNA
PT fragment
PS Disclosure; Fig 13; 67pp; Japanese.
CC This protein is encoded by ORF1 of the 23666bp sequence
CC isolated from L.lactamgenus and comprising the genes for the
CC cephalosporin biosynthetic enzymes listed in the KEYWORDS. Plasmids
CC containing at least one of ORF's 1-9 can be used to transform
CC microbes, such as bacteria or yeast.
CC See also Q10191-2.
SQ Sequence 3722 AA;

Query Match 46.3%; Score 63; DB 2: Length 3722;
Best Local Similarity 50.0%; Pred.No. 4.19e+01;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 1159 hkl1faigspvtwvg 1174
QY 43 HTVLFEPGSSSVWVG 58

RESULT 2
 ID R42838 standard; Protein; 165 AA.
 AC R42838;
 DT 10-MAY-1994 (first entry)
 DE Prochymosin N-terminal fragment.
 KW Prochymosin; peptide; aspartic acid; phenylalanine; fusion; peptide;
 KM plasmid; transformation; aspartylphenylalanine; DF.
 OS Synthetic.
 FH Key
 FT Location/Qualifiers
 FT 1..46
 FT /note= "Signal peptide"
 FT 47..165
 FT /note= "Mature protein fragment"
 FT
 PN J05244960-A.
 PD 24-SEP-1993.
 PF 18-FEB-1992; 030821.
 PR 18-FEB-1992; JP-030821.
 PA (AJIN) AJINOMOTO KK.
 DR WPI; 93-338926/43.
 DR N-PSDB; Q49459.
 DR Prochymosin and aspartyl-phenylalanine polymer fusion protein and
 PT DNA - for prodn. of protein using microorganism.
 PS Claim 2: Page 7-8; 1pp; Japanese.
 CC This sequence represents a fragment of prochymosin. The DNA
 CC encoding this sequence may be linked to a further DNA fragment
 CC encoding a peptide in which aspartic acid and phenylalanine are
 CC repeated mutually at the C-terminal of the peptide. The fusion
 CC sequence may be included in a plasmid which is used to transform a
 CC microorganism. This allows preparation of aspartylphenylalanine (DF)
 CC efficiently.
 SQ Sequence 165 AA;
 Query Match 45.6%; Score 62; DB 8; Length 165;
 Best Local Similarity 64.3%; Pred. No. 5.20e+01;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
 Db 76 tvlf-dtgssdfwv 88
 |||| : ||| ||
 QY 44 TVLFHEPGSSSVWV 57

RESULT 3
 ID P80846 standard; protein; 276 AA.
 AC P80846;
 DT 10-SEP-1990 (first entry)
 DE Sequence of proteinase K
 KW Serine protease TW3; proteinase K.
 PN WO-8807381-A.
 PF 06-OCT-1988.
 PR 28-MAR-1988; U01040.
 PR 03-MAR-1987; US-035816.
 PA (AMGE) Amgen Inc.
 PI Samal BB, Stabinsky Y;
 DR WPI; 88-292865/41.
 DR New purified serine protease -
 PT isolated from culture medium of fungus Triticarium album or
 PT produced by recombinant DNA techniques
 PS Disclosure: Fig 9a-9c; 74pp; English.
 CC Fig. 9 represents a comparison of the AA sequences of protease TW3 with
 CC those of proteinase K, subtilisin novo, subtilisin Carlsberg, subtilisin
 CC DY and thermolysin. The AA sequence of the mature TW3 protein
 CC as determined from the nucleotide sequence has approximately 90%
 CC homology with that of proteinase K. There are certain positions where
 CC the TW3 AA sequence resembles that in subtilisins, but not to proteinase
 CC K. For example, at positions 143, a methionine residue occurs in all
 CC subtilisins as well as in protease TW3, while a leucine residue is
 CC present at that position in proteinase K. Similarly at position 219, an
 CC alanine residue is present in protease TW3 and subtilisins, but not in
 CC proteinase K. In addition, the AA fragment, Ser-Thr-, is absent from
 CC proteinase K while being present in the other enzymes (see p80847-50 and
 CC p80851) at position 226 and 227. Purified serine protease (SP), can be
 CC used in detergents and cleansers or spot removers, as a depilatory in
 CC tanning and also in the food industry for the prepn. of protein

CC hydrolysates and in serology for the detection of incomplete antibodies.
 SQ Sequence 276 AA;
 Query Match 45.6%; Score 62; DB 1; Length 276;
 Best Local Similarity 47.1%; Pred. No. 5.20e+01;
 Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;
 Db 197 svldifpgpdtslwigg 213
 :|| : || : || : ||
 QY 44 TVLFHEPGSSSVWVG 59

RESULT 4
 ID P30603 standard; Protein; 365 AA.
 AC P30603;
 DT 14-JUN-1992 (first entry)
 DE Sequence encoded by prorennin cDNA in PCR 10001.
 KW Rennin; renin; enzyme; protease.
 OS Bos taurus.
 PN EP--73029-A.
 PD 02-MAR-1983.
 PF 19-AUG-1982; 107601.
 PR 24-AUG-1981; JP-131631.
 PA (BEPP) BEPPU T.
 PI Beppu T, Uozumi T, Nishimori K;
 DR WPI; 83-22976K/10.
 DR N-PSDB; N30063.
 PT Plasmid contg. calf pro:rennin DNA - and transformed
 PT microorganisms
 PS Example: Page 20-23; 32pp; English.
 CC The inventors claim recombinant plasmids contg. the cDNA of calf
 CC prorennin. Specified plasmids are PCR 10001 and PCR2001 (contg. the
 CC whole sequence plus the lac promoter region). Also new are
 CC microorganisms transformed with the plasmids, esp. E. coli CRL (ATCC
 CC 391710) contg. plasmid PCR2001.
 SQ Sequence 365 AA;
 Query Match 45.6%; Score 62; DB 4; Length 365;
 Best Local Similarity 64.3%; Pred. No. 5.20e+01;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
 Db 72 tvlf-dtgssdfwv 84
 |||| : ||| ||
 QY 44 TVLFHEPGSSSVWV 57

RESULT 5
 ID R05080 standard; protein; 365 AA.
 AC R05080;
 DT 05-OCT-1990 (first entry)
 DE Sequence of calf pro-rennin.
 KW pro-rennin; ds.
 OS Bos taurus.
 PN J02109984-A.
 PD 23-APR-1990.
 PR 01-JAN-1988; 302176.
 PR 01-JAN-1988; JP-302176.
 PA (VEEP) Veppu T.
 DR WPI; 90-188358/22.
 DR N-PSDB; Q04683.
 PT Complex plasmid and microbe - contains calf pro-rennin cDNA.
 PS Disclosure; 32; 13pp; Japanese.
 CC Protein product may be expressed in E.coli expression system from
 CC plasmid pBR322.
 CC Sequence 365 AA;
 Query Match 45.6%; Score 62; DB 1; Length 365;
 Best Local Similarity 64.3%; Pred. No. 5.20e+01;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
 Db 72 tvlf-dtgssdfwv 84
 |||| : |||| ||
 QY 44 TVLFHEPGSSSVWV 57

RESULT 6
 ID P94144 standard; protein: 365 AA.
 AC P94144;
 DT 07-JUN-1990 (first entry)
 DE Prochymosin.
 KW Lactic acid bacteria; cheese; Streptococcus cremoris SK112;
 KW proteinase: pSK112; chymosin; prochymosin.
 PN NL8701378-A.
 PD 02-JAN-1989.
 PF 12-JUN-1987; 001378.
 PR 12-JUN-1987; NL-001378.
 PA (NEZU-) Nederl Ins Zuivelon.
 PI Simons AFM, De Vos WM;
 DR WPI: 89-030097/04.
 DR N-PSDB: N91157.
 DT DNA fragment having region specific for lactic acid bacteria -
 PT is contained in plasmid in microorganism used in prodn. of
 PT protein and food prodn. eg cheese.
 PS Disclosure: p: Dutch.
 CC The DNA encoding prochymosin can be cloned into a plasmid (esp. from
 CC S. cremoris SK112) and used to produce large ants of the protein by
 CC recombinant DNA techniques. This could overcome shortages of prochymosin
 CC due to a shortage of calf stomachs and increasing cheese prodn.
 CC Prochymosin is also used in prodn. of yoghurt, butter and buttermilk.
 CC See also P94145 and P94146.
 SQ Sequence 365 AA;

Query Match 45.6%; Score 62; DB 1; Length 365;
 Best Local Similarity 64.3%; Pred. No. 5.20e+01;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 72 tvlf-dtgssdfwv 84
 |||| : ||| ||
 Qy 44 TVLFHEPGSSSSVWV 57

RESULT 7
 ID P40078 standard; Protein: 375 AA.
 AC P40078;
 DT 02-FEB-1992 (first entry)
 DE Sequence encoded by prochymosin gene.
 KW Prochymosin expression vector; E.coli trp operon; chymosin.
 OS Bos taurus.
 PN EP-121775-A.
 PD 17-OCT-1984.
 PF 07-MAR-1984; 102451.
 PR 09-MAR-1983; JP-038439.
 PA (BEPPU) BEPPU T.
 PI Beppu T, Uozumi T, Nishimori K, Shimizu N, Kawaguchi Y,
 PI Hidaka M;
 DR WPI: 84-258001/42.
 DR N-PSDB: N40055.
 PT Expression plasmid comprising prochymosin gene and vector -
 PT useful for transforming Escherichia coli for prochymosin prodn.
 PS Disclosure: Fig 1; 59pp; English.
 CC The inventors claim the prochymosin gene comprising a nucleotide
 CC sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or
 CC the 5th codon (CGG) to the 365th codon (ATC); and recombinant
 CC plasmids harboured by Escherichia coli strains deposited as FERM BP-
 CC 262, -263 and -264. Any portion of the nucleotide sequence as
 CC described in N40055 can be used. Also claimed is a vector derived
 CC from plasmid pBR322. Typically plasmid pCR501 is obt'd. from pOCT 2.
 CC The transcriptional direction of pOCT 3 is opposite to that of
 CC pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter
 CC clockwise in pOCT 3.
 SQ Sequence 375 AA;

Query Match 45.6%; Score 62; DB 4; Length 375;
 Best Local Similarity 64.3%; Pred. No. 5.20e+01;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 82 tvlf-dtgssdfwv 94
 |||| : ||| ||
 Qy 44 TVLFHEPGSSSSVWV 57

RESULT 8
 ID P30013 standard; peptide; 379 AA.
 AC P30013;
 DT 25-APR-1992 (first entry)
 DE Sequence encoded by veal chymosin gene.
 KW Protolytic enzyme; zymogen; rennin; chymosin; cheese making;
 KW microbial vector.
 OS Bos taurus.
 PN BE-897201-A.
 PD 03-NOV-1983.
 PF 30-JUN-1983; 017731.
 PR 01-JUL-1982; US-394433.
 PR 13-APR-1983; US-484539.
 PA (GENE-) GENEX CORP.
 DR WPI: 83-820813/47.
 DR N-PSDB: N30022.
 PT Isolated chymosin or rennin and prochymosin genes - plasmid(s)
 PT which replicate in prokaryotic organisms, esp. Escherichia coli,
 PT and organisms used for chymosin biosynthesis
 PS Disclosure: Page 33-36; 43pp; French.
 CC The inventors claim isolated chymosin (rennin) and prochymosin genes
 CC from calves, and plasmids contg. the genes which are capable of
 CC replicating in a prokaryotic organism. The prokaryotic organism is
 CC pref. an Escherichia species, esp. E. coli p Gx 1225 (NRRL B-15061).
 CC The microorganisms transformed by the plasmid are also claimed.
 SQ Sequence 379 AA;

Query Match 45.6%; Score 62; DB 4; Length 379;
 Best Local Similarity 64.3%; Pred. No. 5.20e+01;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 86 tvlf-dtgssdfwv 98
 |||| : ||| ||
 Qy 44 TVLFHEPGSSSSVWV 57

RESULT 9
 ID P20038 standard; Protein: 380 AA.
 AC P20038;
 DT 16-DEC-1992 (first entry)
 DE Pre-prorennin-A protein sequence.
 KW Pre-pro-rennin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;
 KW protease; milk-clotting enzyme; ss.
 OS Bos taurus.
 PN GB2091271-A.
 PD 28-JUL-1982.
 PF 15-JAN-1982; 001120.
 PR 16-JAN-1981; US-225717.
 PR 01-DEC-1981; US-325481.
 PA (COLB) COLLABORATIVE RES INC.
 PI Alford BL, Mao J, Moir DT;
 DR WPI: 82-62028E/30 (62028E).
 DR P-PSDB: P20038.
 PT Transformed cells producing rennin and its precursors - contg.
 PT appropriate recombinant DNA material
 PS Disclosure; Table 1; 39pp; English.
 CC DNA sequences either side of the protein sequence
 CC can be removed and are not essential to use of the gene in
 CC expression. The protein may be expressed in E. coli using
 CC plasmid pCGE31. The resulting expressed enzyme is a well
 CC known milk-clotting enzyme used in cheese-making.
 SQ Sequence 380 AA;

Query Match 45.6%; Score 62; DB 5; Length 380;
 Best Local Similarity 64.3%; Pred. No. 5.20e+01;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 87 tvlf-dtgssdfwv 99

QY 44 TVLFHEPGSSVWV 57
|||| : ||| ||

RESULT 10
ID P40559 standard; Protein; 381 AA.
AC P40559;
DT 04-FEB-1992 (first entry)
DE Sequence of a polypeptide displaying milk clotting activity.
KW Cheese-making; recombinant protein; rennet substitute; milk clot.
OS Bos taurus.
FH Key Location/Qualifiers
FT peptide 1..16
FT /label= signal
FT 17..381
FT protein
FN EP-123928-A.
PD 07-NOV-1984.
PF 30-MAR-1984; 103551.
PR 31-MAR-1983; US-480860.
PA (CODON) CODON GENETIC ENG.
PI Cashion LM, McCaman MT, Rice CW, Sias SR;
DR WPI: 84-277277/45.
DR N-PSDB: N40295.
PT Recombinant DNA coding for milk clotting polypeptide - which is
PT expressed in transformed bacteria
PS Claim 10; Fig 2; 39pp; English.
CC Also claimed is E.coli JM83/pLC7 (ATCC 39325) which is transformed
CC with pLC7 contg. the prorennin derived sequence fused in phase with
CC B-galactosidase. The pLC7 prorennin expression plasmid includes
CC sequences which code for both the pseudorennin and mature rennin
CC cleavage sites between AAs 28-29 and AAs 42-43, respectively.
SQ Sequence 381 AA;

Query Match 45.6%; Score 62; DB 4; Length 381;

Best Local Similarity 64.3%; Pred. No. 5.20e+01;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 88 tvlf-dtgssdfwv 100
|||| : ||| ||

QY 44 TVLFHEPGSSVWV 57

RESULT 11
ID P30086 standard; Protein; 381 AA.
AC P30086;
DT 14-JUN-1992 (first entry)
DE Sequence encoded by cDNA sequence corresponding to one of the
DE allelic forms (B) of bovine preprochymosin.
KW Chymosin; enzyme; rennet; cheese.
OS Bos taurus.
FH Key Location/Qualifiers
FT peptide 1..16
FT /label= signal
FN EP--77109-A.
PD 20-APR-1983.
PF 13-OCT-1982; 201272.
PR 14-OCT-1981; GB-031004.
PA (UNIL) UNILEVER NV.
PI Maat J, Verrips CT, Ledebor AM, Edens L;
DR WPI: 83-39636K/17.
DR N-PSDB: N30049.
PT DNA molecules comprising genes for preprochymosin - used to
PT transform microorganisms to give strain producing the
PT prepro-enzyme and its allelic and maturation forms
PS Claim 2; Fig 1; 53pp; English.
CC Preprochymosin is an intermediate (via prochymosin and
CC pseudochymosin) for the enzyme chymosin, which is the essential milk-
CC clotting component of rennet and is used in cheese manufacture.
CC N30049 corresp. to mRNA isolated from the fourth stomach of a
CC pre-ruminant calf (abomasum, Frisian cow).
SQ Sequence 381 AA;

Query Match 45.6%; Score 62; DB 4; Length 381;

Best Local Similarity 64.3%; Pred. No. 5.20e-01;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 88 tvlf-dtgssdfwv 100
|||| : ||| ||

QY 44 TVLFHEPGSSVWV 57

RESULT 12
ID P30446 standard; Protein; 381 AA.
AC P30446;
DT 03-AUG-1992 (first entry)
DE Sequence encoded by preprochymosin cDNA.
KW Milk-clotting; cheese making; enzyme; zymogen.
OS Cow.
FH Key Location/Qualifiers
FT region 1..16
FT /label= preprochymosin
FT region 17..58
FT /label= prochymosin
FT region 59..381
FT /label= chymosin
FN GB2100737-A.
PD 06-JAN-1983.
PF 11-JUN-1982; 017096.
PR 10-FEB-1982; GB-003907.
PA (CELL-) CELLTech LTD.
PI Carey NH, Harris TJR, Lowe PA, Doel MT, Emtage JS;
DR WPI: 83-00545K/01.
DR N-PSDB: N30209.
PT Prodn. of calf stomach chymosin for cheese making - by
PT cultivation of micro-organisms transformed with vector system
PS Claim 41; Fig 4; 26pp; English.
CC The inventors claim a method for the prodn. of calf stomach chymosin
CC for cheese making. Genes and polypeptides for preprochymosin, a
CC prochymosin and chymosin are claimed, as are vector systems and a
CC prochymosin primer.
SQ Sequence 381 AA;

Query Match 45.6%; Score 62; DB 4; Length 381;

Best Local Similarity 64.3%; Pred. No. 5.20e-01;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 88 tvlf-dtgssdfwv 100
|||| : ||| ||

QY 44 TVLFHEPGSSVWV 57

RESULT 13
ID R20730 standard; Protein; 381 AA.
AC R20730;
DT 13-MAY-1992 (first entry)
DE Prochymosin (prorennin).
KW Rennet; zymogen; autocatalytic activation; pseudorennin; milk;
KW clotting activity; renin A, renin B.
OS Bos taurus.
FH Key Location/Qualifiers
FT protein 1..381
FT /label= preprorennin
FT /note= "also known as preprochymosin"
FT peptide 1..17
FT /label= signal_peptide
FT cleavage_site 43..44
FT /note= "autocatalytic cleavage at pH 2"
FT cleavage_site 59..60
FT /note= "autocatalytic cleavage pH 4.7"
FT misc_difference 306
FT /note= "amino acid determining A or B form
FT of rennin"
FT misc_difference 218
FT /note= "Asp in published sequence"
FN US5082775-A.
PD 21-JAN-1992.

PF 28-OCT-1988: 263927.
 PR 31-MAR-1983: US-480860.
 PR 11-MAY-1984: US-609495.
 PR 28-APR-1986: US-856700.
 PR 12-DEC-1986: US-940199.
 PR 28-OCT-1988: US-263927.
 PA (BERL-) BERLEX LABS INC.
 PI McCaman MT, King JF;
 DR WPI: 92-049149/06.
 DR N-PSDB: Q20949.
 PT Isolating heterologous polypeptide from bacterial inclusion
 PT bodies - by lysing cells, extn. with nonionic detergent and sepg.
 PT Insoluble polypeptide
 PS Disclosure: Fig 6: 21pp: English.
 CC The prorennin (prochymosin) sequence was deduced from the DNA
 CC sequence obt'd. by screening a cDNA library with rennin specific
 CC probes derived from the published amino acid sequence of rennin
 CC (Foltmann et al., J. Biol. Chem. 254, 8447-8456 (1979)) (see 223291.2).
 CC The deduced protein sequence was the same as the published sequence
 CC after cleavage of the signal peptide, except for an amino acid change
 CC at residue 218, which may be an artefact of protein sequencing in
 CC the original sequence. The cleavage product of prorennin, rennin is
 CC synthesised in two active forms rennin A, and rennin B. The cleaved
 CC secreted protein undergoes activation in the acidic conditions of the
 CC stomach to generate mature rennin by two autocatalytic cleavage events.
 CC Rennin is an active component of rennet which is used to clot milk in
 CC the process of making cheese.
 CC See also R22417.
 SQ Sequence 381 AA;

Query Match 45.6%; Score 62; DB 4; Length 381;
 Best Local Similarity 64.3%; Pred. No. 5.20e+01;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 88 tvlf-dtgssdfwv 100
 |||| : ||| ||
 QY 44 TVLFHEPGSSSVWV 57

RESULT 14
 ID P40218 standard; Protein; 381 AA.
 AC P40218;
 DT 25-JAN-1992 (first entry)
 DE Sequence of rennin encoded by recombinant CGF4
 KW Yeast expression vector; GAL1 promoter; Saccharomyces cerevisiae.
 OS Bos taurus.
 PN GB2137208-A.
 PD 03-OCT-1984.
 PF 28-FEB-1984: 405129.
 PR 28-FEB-1983: US-470911.
 PA (COLB) COLLABORATIVE RES INC.
 PI Botstein D, Davis RW, Fink GR, Taunton-Rigby A, Knowlton RG,
 PI Mao JI, Moir JT, Goff CG;
 DR WPI: 84-245517/40.
 DR N-PSDB: N40180.
 PT DNA segment contg. GAL1 promoter linked to gene - useful for
 PT direction of expression of the gene in yeast cell
 PT Example: Table 4, Page 21-23: 35pp; English.
 PS The inventors claim a DNA segment contg. GAL1 promoter linked to
 CC gene - useful for direction of expression of the gene in yeast cell.
 CC The recombinant material carrying a GAL1 promoter of the yeast
 CC galactokinase gene may be used in expressing a desired protein, esp.
 CC bovine growth hormone, interferon, prorennin or preprorennin, in the
 CC yeast cell. Strains of Saccharomyces cerevisiae producing the
 CC polypeptides are produced. Yeast strains deposited as ATCC 20643,
 CC 20661, 20662 and 20663, strain designations CGY 196, 457, 461 and
 CC 528, resp. are new.
 SQ Sequence 381 AA;

Query Match 45.6%; Score 62; DB 4; Length 381;
 Best Local Similarity 64.3%; Pred. No. 5.20e+01;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 88 tvlf-dtgssdfwv 100
 |||| : ||| ||
 QY 44 TVLFHEPGSSSVWV 57

RESULT 15

ID P94370 standard; protein; 450 AA.
 AC P94370;
 DT 10-MAR-1993 (revised)
 DE Sequence encoded by BamHI insert from pDM100PC.
 DE Kluyveromyces; pDM100PC; chymosin; tissue plasminogen activator.
 KW EP-301670-A.
 PN 01-JAN-1989.
 PD 28-JUL-1988: 201632.
 PF 28-JUL-1987: US-078539.
 PR (KONN) Gist-Brocades NV.
 PA van den Berg JA, van Ooyen AJJ, Rietveld K;
 DR WPI: 89-033565/05.
 DR N-PSDB: N91185.
 PT Kluyveromyces host cells for producing polypeptide(s) -
 PT used for highly efficient prodn. of eg chymosin tissue
 PT plasminogen activator or human serum albumin.
 PS Disclosure: 56 pp; English.
 CC The pDM100PC sequence product comprises a fusion peptide of the alpha-
 CC factor of S.cerevisiae and prochymosin and transcriptional and regulatory
 CC control regions. The Kluyveromyces expression systems provide highly
 CC efficient secretion and processing of a wide variety of proteins.
 CC Sequences identical to those published in EP301669.
 SQ Sequence 450 AA;

Query Match 45.6%; Score 62; DB 1; Length 450;
 Best Local Similarity 64.3%; Pred. No. 5.20e+01;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 157 tvlf-dtgssdfwv 169
 |||| : ||| ||
 QY 44 TVLFHEPGSSSVWV 57

Search completed: Thu Jul 8 18:12:03 1999
 Job time : 18 secs.

THIS PAGE BLANK (USPTO)

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:11:15 1999; MasPar time 5.15 Seconds
Tabular output not generated. 140.146 Million cell updates/sec

Title: >US-09-041-236-2
Description: (42-59) from US09041236.pep (4 of 45)
Perfect Score: 136
Sequence: 1 PHTVLFHEPGSSVWVG 18

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 27.933; Variance 44.714; scale 0.625

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	88	64.7	653	2	T03102 semaphorin homolog A3	1.31e-04
2	67	49.3	388	2	pepsin A (EC 3.4.23.1)	8.42e-01
3	67	49.3	388	2	pepsin A (EC 3.4.23.1)	8.42e-01
4	66	48.5	439	2	hypothetical protein	1.24e+00
5	65	47.8	182	2	hypothetical protein	1.82e+00
6	65	47.8	926	1	iodide peroxidase (EC 1.8.2e+00)	1.82e+00
7	64	47.1	392	1	gastricsin (EC 3.4.23.4)	2.65e+00
8	63	46.3	424	2	hypothetical protein	3.86e+00
9	62	45.6	130	2	hypothetical protein	5.60e+00
10	62	45.6	377	1	gastricsin (EC 3.4.23.4)	5.60e+00
11	62	45.6	381	1	chymosin (EC 3.4.23.4)	5.60e+00
12	62	45.6	381	1	chymosin (EC 3.4.23.4)	5.60e+00
13	62	45.6	384	2	gastricsin (EC 3.4.23.4)	5.60e+00
14	62	45.6	388	2	gastricsin (EC 3.4.23.4)	5.60e+00
15	62	45.6	389	3	pepsinogen C - Chiocke	5.60e+00
16	62	45.6	394	2	gastricsin (EC 3.4.23.4)	5.60e+00
17	61	44.9	52	2	cathepsin E (EC 3.4.23.4)	8.10e+00
18	61	44.9	110	1	pepsin A (EC 3.4.23.1)	8.10e+00
19	61	44.9	249	1	phosphatidate cytidyl	8.10e+00
20	61	44.9	365	2	cathepsin E (EC 3.4.23.4)	8.10e+00
21	61	44.9	382	3	pepsinogen A - Chiocke	8.10e+00
22	61	44.9	385	1	pepsin A (EC 3.4.23.1)	8.10e+00
23	61	44.9	387	2	pepsin (EC 3.4.23.1)	8.10e+00

24	61	44.9	387	2	C38302	pepsin (EC 3.4.23.1)	8.10e+00
25	61	44.9	387	2	D38302	pepsin (EC 3.4.23.1)	8.10e+00
26	61	44.9	387	2	E38302	pepsin (EC 3.4.23.1)	8.10e+00
27	61	44.9	388	1	S19684	pepsin A (EC 3.4.23.1)	8.10e+00
28	61	44.9	388	1	PEMQAR	pepsin A (EC 3.4.23.1)	8.10e+00
29	61	44.9	388	1	PEMQAJ	pepsin A (EC 3.4.23.1)	8.10e+00
30	61	44.9	391	2	A43356	cathepsin E (EC 3.4.23.1)	8.10e+00
31	61	44.9	396	2	A34401	cathepsin E (EC 3.4.23.1)	8.10e+00
32	61	44.9	398	2	S66465	cathepsin E (EC 3.4.23.1)	8.10e+00
33	61	44.9	406	1	REHUK	renin (EC 3.4.23.15)	8.10e+00
34	61	44.9	428	2	S47096	cynarase (EC 3.4.23.1)	8.10e+00
35	61	44.9	442	2	JC5077	aspartic proteinase (EC 3.4.23.1)	8.10e+00
36	61	44.9	496	2	JS0732	aspartic proteinase (EC 3.4.23.1)	8.10e+00
37	61	44.9	508	2	S19697	aspartic proteinase (EC 3.4.23.1)	8.10e+00
38	61	44.9	509	2	S66516	oryzasin (EC 3.4.23.1)	8.10e+00
39	61	44.9	509	2	S49349	cyprosin (EC 3.4.23.1)	8.10e+00
40	61	44.9	1124	2	D65032	hypothetical protein	8.10e+00
41	61	44.9	1357	2	S61187	probable membrane protein	8.10e+00
42	61	44.9	2334	2	S32920	cell wall-associated	8.10e+00
43	60	44.1	383	2	A41443	pepsin (EC 3.4.23.1)	1.16e+01
44	60	44.1	388	1	S19682	pepsin A (EC 3.4.23.1)	1.16e+01
45	60	44.1	694	2	F70868	hypothetical glycine-	1.16e+01

ALIGNMENTS

RESULT 1
ENTRY T03102 #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS T03102
REFERENCE Z14840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type DNA
##residues 1-653 #label ENS
##cross-references EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY #length 653 #molecular_weight 73645 #checksum 5501
Query Match 64.7%; Score 88; DB 2; Length 653;
Best Local Similarity 72.2%; Pred. No. 1.31e-04;
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 96 PHTVLFHSLNSSDYVVG 113
Qy 42 PHTVLFHEPGSSVWVG 59
RESULT 2
ENTRY A30142 #type complete
TITLE pepsin A (EC 3.4.23.1) 5 precursor - human
ALTERNATE_NAMES pepsinogen 5
ORGANISM #formal_name Homo sapiens #common_name man
DATE 07-Jun-1990 #sequence_revision 20-Aug-1994 #text_change 29-Aug-1997
ACCESSIONS A30142; S02664; S02542; PX0027; PX0025; PX0026; A22434
REFERENCE A91627
#authors Evers, M.P.J.; Zelle, B.; Bebelman, J.P.; van Beusechem, V.; Kraakman, L.; Hoffer, M.J.V.; Pronk, J.C.; Mager, W.H.; Plant, R.J.; Eriksson, A.W.; Frants, R.R.
#journal Genomics (1989) 4:232-239
#title Nucleotide sequence comparison of five human pepsinogen A (PGA) genes: evolution of the PGA multigene family.
#cross-references MIM:89233110
#accession A30142
##molecule_type DNA
##residues 1-27, 'F', 29-388 #label EVE
##cross-references GB:M26025

```

#note      the authors translated the codon TTC for residue 28 as
           leu, GGC for residue 36 as Glu, GAG for residue 376 as
           Asp, and CAG for residue 381 as Gly

REFERENCE  S02663
#authors   Foltmann, B.
#journal   FEBS Lett. (1988) 241:69-72
#title     Activation of human pepsinogens.
#cross-references MUID:89065108
#accession S02664
#molecule_type protein
#residues  16-68 #label FOL

REFERENCE  S02542
#authors   Bank, R.A.; Crusius, B.C.; Zwiers, T.; Meuwissen, S.G.M.;
           Arwert, F.; Pronk, J.C.
#journal   FEBS Lett. (1988) 238:105-108
#title     Identification of a Glu > Lys substitution in the activation
           segment of human pepsinogen A-3 and -5 isozymogens by
           peptide mapping using endoproteinase Lys-C.
#cross-references MUID:89005849
#accession S02542
#status    preliminary
#molecule_type protein
#residues  16-58 #label BAN

REFERENCE  PX0023
#authors   Ahtauda, S.B.P.; Tanji, M.; Kageyama, T.; Takahashi, K.
#journal   J. Biochem. (1989) 106:920-927
#title     A comparative study on the NH2-terminal amino acid sequences
           and some other properties of six isozymic forms of human
           pepsinogens and pepsins.
#cross-references MUID:90130402
#accession PX0027
#molecule_type protein
#residues  16-100 #label ATH

#accession PX0025
#molecule_type protein
#residues  16-60 #label AT2

#accession PX0026
#molecule_type protein
#residues  16-100 #label AT3

REFERENCE  A22434
#authors   Ichihara, Y.; Sogawa, K.; Takahashi, K.
#journal   J. Biochem. (1985) 98:483-492
#title     Isolation of human, swine, and rat prepepsinogens and calf
           preprochymosin, and determination of the primary structures
           of their NH2-terminal signal sequences.
#cross-references MUID:86059312
#accession A22434
#molecule_type protein
#residues  1-15,'XXX',19-20,'X',22,'XX',25-26,'X',28 #label ICH

GENETICS
#gene      GDB:PGA5
#cross-references GDB:119484; OMIM:169730
#map_position 11q13-11q13
#introns    19/1: 73/3; 113/1: 152/3; 219/2: 258/2; 306/3: 339/3
CLASSIFICATION
#superfamily pepsin
KEYWORDS    aspartic proteinase; hydrolase; phosphoprotein; protein
           digestion
FEATURE
1-15        #domain signal sequence #status experimental #label SIG\
16-59       #domain activation peptide #status experimental #label
           APT\
63-388      #product pepsin A 4 #status predicted #label MAT\
94,277      #active_site Asp #status predicted\
107-112,268-272,
311-344     #disulfide_bonds #status predicted\
130         #binding_site phosphate (Ser) (covalent) #status
           predicted
SUMMARY     #length 388 #molecular-weight 42004 #checksum 8564

Query Match 49.3%; Score 67; DB 2; Length 388;
Best Local Similarity 64.3%; Pred. No. 8.42e-01;
Matches 9; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 90 TVLFD-TGSSNLWV 102

```

```

|||| : |||::||
QY 44 TVLFHEPGSSVWV 57

RESULT 3
ENTRY   B30142 #type complete
TITLE   pepsin A (EC 3.4.23.1) 4 precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     07-Jun-1990 #sequence_revision 20-Aug-1994 #text_change
29-Aug-1997
B30142; E22434
A91627
Evers, M.P.J.; Zelle, B.; Bebelman, J.P.; van Beusechem, V.;
Kraakman, L.; Hoffer, M.J.V.; Pronk, J.C.; Mager, W.H.;
Planta, R.J.; Eriksson, A.W.; Frants, R.R.
Genomics (1989) 4:232-239
#journal Nucleotide sequence comparison of five human pepsinogen A
#title   (PGA) genes: evolution of the PGA multigene family.
#cross-references MUID:89233110
#accession B30142
#molecule_type DNA
#residues  1-27,'F',29-388 #label EVE
#note      the authors translated the codon TTC for residue 28 as
           Leu, GGC for residue 36 as Glu, GAG for residue 376 as
           Asp, and CAG for residue 381 as Gly

REFERENCE  A22434
#authors   Ichihara, Y.; Sogawa, K.; Takahashi, K.
#journal   J. Biochem. (1985) 98:483-492
#title     Isolation of human, swine, and rat prepepsinogens and calf
           preprochymosin, and determination of the primary structures
           of their NH2-terminal signal sequences.
#cross-references MUID:86059312
#accession E22434
#molecule_type protein
#residues  1-15,'XXX',19-20,'X',22,'XX',25-26,'X',28 #label ICH

GENETICS
#gene      GDB:PGA4
#cross-references GDB:119483; OMIM:169720
#map_position 11q13-11q13
#introns    19/1: 73/3; 113/1: 152/3; 219/2: 258/2; 306/3: 339/3
CLASSIFICATION
#superfamily pepsin
KEYWORDS    aspartic proteinase; hydrolase; phosphoprotein; protein
           digestion; zymogen
FEATURE
1-15        #domain signal sequence #status experimental #label SIG\
16-59       #domain activation peptide #status experimental #label
           APT\
63-388      #product pepsin A 4 #status predicted #label MAT\
94,277      #active_site Asp #status predicted\
107-112,268-272,
311-344     #disulfide_bonds #status predicted\
130         #binding_site phosphate (Ser) (covalent) #status
           predicted
SUMMARY     #length 388 #molecular-weight 42021 #checksum 9065

Query Match 49.3%; Score 67; DB 2; Length 388;
Best Local Similarity 64.3%; Pred. No. 8.42e-01;
Matches 9; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 90 TVLFD-TGSSNLWV 102
|||| : |||::||
QY 44 TVLFHEPGSSVWV 57

RESULT 4
ENTRY   S69582 #type complete
TITLE   hypothetical protein YDR527w - yeast (Saccharomyces
           cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change
21-Nov-1997
#accession S69582
REFERENCE  S69553

```

#authors Dietrich, F.S.
#submission submitted to the EMBL Data Library, August 1995
#description The sequence of S. cerevisiae cosmid 8166, 9787, 9717, and
lambda 3073.
#accession S69382
##molecule_type DNA
##residues 1-439 ##label DIE
##cross-references EMBL:U33057; NID:g927764; PID:g927794; MIPS:YDR527w

GENETICS
#map_position 4R
SUMMARY #length 439 #molecular-weight 50106 #checksum 9410
Query Match 48.5%; Score 66; DB 2; Length 439;
Best Local Similarity 55.6%; Pred. No. 1.24e+00;
Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Db 129 NTPFAIEGASGTWVG 146
: | | | | | | | | | |
Qy 43 HTVLFHE-PGSSSVWVG 59

RESULT 5
ENTRY #type complete
TITLE hypothetical protein Rv0513 - Mycobacterium tuberculosis
(strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998

ACCESSIONS
REFERENCE F70509
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squares, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Squares, R.; Sulston, J.E.;
Nature (1998) 393:537-544
#journal Deciphering the biology of Mycobacterium tuberculosis from
#title the complete genome sequence.
#cross-references MUID:98295987
#accession F70509
##status preliminary; nucleic acid sequence not shown;
translation not shown

##molecule_type DNA
##residues 1-182 ##label COL
##cross-references GB:297831; GB:AL123456; NID:g3261825; PID:e329664;
PID:g2266716
##experimental_source strain H37Rv

GENETICS
#gene Rv0513
SUMMARY #length 182 #molecular-weight 19401 #checksum 4145
Query Match 47.8%; Score 65; DB 2; Length 182;
Best Local Similarity 61.5%; Pred. No. 1.82e+00;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 11 LLFEPGASWYV 23
: | | | | | | | | | |
Qy 45 VLFEPGSSVWV 57

RESULT 6
ENTRY #type complete
TITLE iodide peroxidase (EC 1.11.1.8) precursor, thyroid - pig
ALTERNATE_NAMES thyroid peroxidase
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
22-Jan-1999
ACCESSIONS A27416; A24632; S29051

A27416
Magnusson, R.P.; Gestautas, J.; Taurog, A.; Rapoport, B.
J. Biol. Chem. (1987) 262:13885-13888
#title Molecular cloning of the structural gene for porcine thyroid
peroxidase.
#cross-references MUID:88007624
#accession A27416
##molecule_type mRNA
##residues 1-926 ##label MAG
##cross-references GB:J03463

REFERENCE A24632
#authors Magnusson, R.P.; Gestautas, J.; Seto, P.; Taurog, A.;
Rapoport, B.
#journal FEBS Lett. (1986) 208:391-396
#title Isolation and characterization of a cDNA clone for porcine
thyroid peroxidase.
#cross-references MUID:87054611
#accession A24632
##molecule_type mRNA
##residues 595-926 ##label MA2
##cross-references GB:X04645; GB:J03463; NID:g2141; PID:g2142

REFERENCE S29051
#authors Rawitch, A.B.; Pollock, G.; Yang, S.X.; Taurog, A.
#journal Arch. Biochem. Biophys. (1992) 297:321-327
#title Thyroid peroxidase glycosylation: the location and nature of
the N-linked oligosaccharide units in porcine thyroid
peroxidase.
#cross-references MUID:92359545
#accession S29051
##molecule_type protein
##residues 110-115;267-274;301-308;340-351 ##label RAW

CLASSIFICATION #superfamily thyroid peroxidase; complement factor H repeat
homology; EGF homology; myeloperoxidase homology
KEYWORDS glycoprotein; heme; oxidoreductase; thyroid gland; thyroid
hormone biosynthesis; transmembrane protein

FEATURE
1-14 #domain signal sequence #status predicted #label SIG\
15-946 #product iodide peroxidase #status predicted #label MAT\
36-730 #domain myeloperoxidase homology #label MPX\
740-792 #domain complement factor H repeat homology #label FHL\
798-836 #domain EGF homology #label EGF\
852-869 #domain transmembrane #status predicted #label TMW\
129,265,277,307, #binding_site carbohydrate (Asn) (covalent) #status
342 predicted

SUMMARY #length 926 #molecular-weight 100442 #checksum 2698
Query Match 47.8%; Score 65; DB 1; Length 926;
Best Local Similarity 42.9%; Pred. No. 1.82e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 623 LYQHPDNIDVWLGG 636
: | | | | | | | | | |
Qy 46 LFEPGSSVWVG 59

RESULT 7
ENTRY #type complete
TITLE gastrin (EC 3.4.23.3) precursor - rat
ALTERNATE_NAMES pepsinogen C
CONTAINS pepsin A (EC 3.4.23.1) precursor
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 30-Jun-1988 #sequence_revision 05-Aug-1994 #text_change
20-Mar-1998

ACCESSIONS A33510; A24608; C22434; A05145; A61298
REFERENCE A33510
#authors Ishihara, T.; Ichihara, Y.; Hayano, T.; Katsura, I.; Sogawa,
K.; Fujii-Kuriyama, Y.; Takahashi, K.
#journal J. Biol. Chem. (1989) 264:10193-10199
#title Primary structure and transcriptional regulation of rat
pepsinogen C gene.
#cross-references MUID:89255508
#accession A33510

AD169)	
ORGANISM	#formal_name human cytomegalovirus, human herpesvirus 5
#note	host Homo sapiens (man)
DATE	07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Sep-1997
ACCESSIONS	S09884
REFERENCE	S09749
#authors	Chee, M.S.; Bankier, A.T.; Beck, S.; Bohnl, R.; Brown, C.M.; Carny, R.; Horsnell, T.; Hutchinson III, C.A.; Kouzarides, T.; Martignetti, J.A.; Preddie, E.; Satchwell, S.C.; Tomlinson, P.; Weston, K.M.; Barrell, B.G.
#journal	Curr. top. Microbiol. Immunol. (1990) 154:125-169
#title	Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169.
#cross-references	MUID:90269039
#accession	S09884
#status	nucleic acid sequence not shown; translation not shown
#molecule_type	DNA
#residues	1-424 #label CHE
#cross-references	EMBL:X17403; NID:g59591; PID:g1780899
#note	this sequence was submitted to the EMBL Data Library, December 1989
SUMMARY	#length 424 #molecular-weight 45463 #checksum 2954
Query Match	46.3%; Score 63; DB 2; Length 424;
Best Local Similarity	38.9%; Pred. No. 3.86e+00;
Matches	7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
Db	352 PRSTSYHETGYQMVVG 369
Qy	1::: : :
	42 PHTVLFEPGSSVWVG 59
RESULT	9
ENTRY	F71035 #type complete
TITLE	hypothetical protein PH1574 - Pyrococcus horikoshii
ORGANISM	#formal_name Pyrococcus horikoshii
DATE	14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
ACCESSIONS	F71035
REFERENCE	A71000
#authors	Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#journal	DNA Res. (1998) 5:55-76
#title	Complete sequence and gene organization of the genome of a hyperthermophilic archaeobacterium, Pyrococcus horikoshii OT3
#cross-references	MUID:98344137
#accession	F71035
#status	preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type	DNA
#residues	1-130 #label KAW
#cross-references	GB:AP000006; NID:g3236133; PID:d1031629; PID:g3258003
#experimental_source	strain OT3
#note	this accession replaces an interim accession for a sequence replaced by GenBank
GENETICS	
#gene	PH1574
SUMMARY	#length 130 #molecular-weight 13603 #checksum 7995
Query Match	45.6%; Score 62; DB 2; Length 130;
Best Local Similarity	53.3%; Pred. No. 5.60e+00;
Matches	8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Db	61 PTTTGGFPNGVTSW 75
Qy	1::: : :
	42 PHTVLFEPGSSVW 56

```

RESULT 10      PEMOCJ      #type fragment
ENTRY          gastricsin (EC 3.4.23.3) precursor - Japanese macaque
TITLE          (fragment)
ALTERNATE_NAMES pepsin C
ORGANISM        #formal_name Macaca fuscata #common_name Japanese macaque
DATE            13-Aug-1986 #sequence_revision 19-Oct-1995 #text_change
                05-Sep-1997

ACCESSIONS     S19683; A00986; A22402; S16066
REFERENCE      S19681
#authors       Kageyama, T.; Tanabe, K.; Koikwai, O.
#journal       Eur. J. Biochem. (1991) 202:205-215
#title         Development-dependent expression of isozymogens of monkey
                pepsinogens and structural differences between them.
#cross-references MUID:92037645
#accession     S19683
#molecule_type mRNA
#residues      1-377 ##label KAG
#cross-references EMBL:X59754; NID:g38072; PID:g38073

REFERENCE      A00986
#authors       Kageyama, T.; Takahashi, K.
#journal       J. Biol. Chem. (1986) 261:4406-4419
#title         The complete amino acid sequence of monkey progastricsin.
#cross-references MUID:86168133
#accession     A00986
#molecule_type protein
#residues      6-330,'V',332-349,'VY',350-377 ##label KA2
#cross-references A22402
#authors       Kageyama, T.; Takahashi, K.
#journal       J. Biochem. (1985) 97:1235-1246
#title         Monkey pepsinogens and pepsins. VII. Analysis of the
                activation process and determination of the NH2-terminal
                60-residue sequence of Japanese monkey progastricsin, and
                molecular evolution of pepsinogens.
#cross-references MUID:85289106
#accession     A22402
#molecule_type protein
#residues      6-65 ##label KA3
COMMENT        This enzyme has more restricted specificity than pepsin A.
COMMENT        The enzyme is activated in a two-step process that gives rise to
                two end products. The shorter, Ser-gastricsin, is the major
                product.
CLASSIFICATION #superfamily pepsin
KEYWORDS        aspartic proteinase; gastric juice; hydrolase; protein
                digestion; stomach

FEATURE
1-5            #domain signal sequence (fragment) #status predicted
               #label SIGV
6-377         #product progastricsin #status experimental #label ZYM\
6-45          #domain activation peptide #status experimental #label
               APT\
46-377        #product Gly-gastricsin #status experimental #label MIN\
49-377        #product Ser-gastricsin #status experimental #label MAT\
31-32         #cleavage_site Phe-Leu (pepsin) #status experimental\
45-46         #cleavage_site Phe-Gly (pepsin) #status experimental\
48-49         #cleavage_site Leu-Ser (pepsin) #status experimental\
80,265        #active_site Asp #status predicted\
93-98,256-260, #disulfide_bonds #status experimental
299-332       #length 377 #checksum 8114

SUMMARY
Query Match      45.6%; Score 62; DB 1; Length 377;
Best Local Similarity 61.5%; Pred. No. 5.60e+00;
Matches          8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db              77 VLF-DTGSNNLWV 88
                ||| : |||::||
QY              45 VLFHEPGSSVWV 57

RESULT 11

```

```

CMBO           #type complete
chymosin (EC 3.4.23.4) precursor - bovine
ALTERNATE_NAMES chymosin; preprorennin; rennin B
CONTAINS         #formal_name Bos primigenius taurus #common_name cattle
ORGANISM         24-Apr-1984 #sequence_revision 09-Sep-1994 #text_change
DATE            26-Feb-1999
ACCESSIONS      A25631; A93419; A44608; A92259; A44620; A91935; A91495;
                D22434; A00985
REFERENCE      A25631
#authors       Hidaka, M.; Sasaki, K.; Uozumi, T.; Beppu, T.
#journal       Gene (1986) 43:197-203
#title         Cloning and structural analysis of the calf prochymosin gene.
#cross-references MUID:86301873
#accession     A25631
#molecule_type DNA
#residues      1-16,'T',18-381 ##label HID
#accession     A93419
#authors       Harris, T.J.R.; Lowe, P.A.; Lyons, A.; Thomas, P.G.; Eaton,
                M.A.W.; Millican, T.A.; Patel, T.P.; Bose, C.C.; Carey,
                N.H.; Doel, M.T.
#journal       Nucleic Acids Res. (1982) 10:2177-2187
#title         Molecular cloning and nucleotide sequence of cDNA coding for
                calf preprochymosin.
#cross-references MUID:82221400
#contents       prochymosin B
#accession     A93419
#molecule_type mRNA
#residues      1-229,'N',231-381 ##label HAR
#cross-references GB:J00003; NID:g162859; PID:g162860

REFERENCE      A44608
#authors       Nishimori, K.; Kawaguchi, Y.; Hidaka, M.; Uozumi, T.; Beppu,
                T.
#journal       J. Biochem. (1982) 91:1085-1088
#title         Nucleotide sequence of calf prorennin cDNA cloned in
                Escherichia coli.
#cross-references MUID:82189915
#accession     A44608
#molecule_type mRNA
#residues      7-14,'X',16-42,'L',44-87,'N',89-301,'D',303-324,'I',
                326-334,'G',336-342,'T',344-379,'T',381 ##label NIS
#cross-references GB:J00004
#note          authors translated the codon TTG for residue 43 as Phe,
                ACC for residue 82 as Tyr, AAC for residue 88 as Thr,
                ATC for residue 325 as Met, ACT for residue 343 as
                Ser, and ACC for residue 380 as Ala

REFERENCE      A92259
#authors       Foltmann, B.; Pedersen, V.B.; Kauffman, D.; Wybrandt, G.
#journal       J. Biol. Chem. (1979) 254:8447-8456
#title         The primary structure of calf chymosin.
#cross-references MUID:79239460
#contents       chymosin B; disulfide bonds
#accession     A92259
#molecule_type protein
#residues      59-217,'D',219-381 ##label FOL
#accession     A44620
#authors       Foltmann, B.; Pedersen, V.B.; Jacobsen, H.; Kauffman, D.;
                Wybrandt, G.
#journal       Proc. Natl. Acad. Sci. U.S.A. (1977) 74:2321-2324
#title         The complete amino acid sequence of prochymosin.
#accession     A44620
#molecule_type protein
#residues      17-173,'T',175-217,'D',219-251,'Y',253-381 ##label FO2
#accession     A91935
#authors       Chang, W.J.; Takahashi, K.
#journal       J. Biochem. (1974) 76:467-474
#title         The structure and function of acid proteases. III. Isolation
                and characterization of the active-site peptides from
                bovine rennin.
#cross-references MUID:75060332
#contents       active site peptides
#accession     A91935
#molecule_type protein

```

##residues 92-96;274-278,280 ##label CHA
##note the authors did not find Lys-279 in their active site peptide

REFERENCE A91495
#authors Moir, D.; Mao, J.; Schumm, J.W.; Vovis, G.F.; Alford, B.L.; Taunton-Rigby, A.
#journal Gene (1982) 19:127-138
#title Molecular cloning and characterization of double-stranded cDNA coding for bovine chymosin.
#cross-references MUID:83054629
#contents prochymosin A
#accession A91495
##molecule_type mRNA
##residues 1-301,'D',303-381 ##label MOI
REFERENCE A22434
#authors Ichihara, Y.; Sogawa, K.; Takahashi, K.
#journal J. Biochem. (1985) 98:483-492
#title Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, and determination of the primary structures of their NH2-terminal signal sequences.
#cross-references MUID:86059312
#accession D22434
##molecule_type protein
##residues 1-18,'XX',21,'XX',24,'X',26,'X',28,'X',30 ##label ICH
COMMENT The sequence of variant B is shown.
COMMENT Chymosin is synthesized in the mucosa of the abomasum (fourth stomach) of young (unweaned) ruminants. The enzyme hydrolyzes casein to paracasein.
COMMENT Forms A and B are probably allelic variants. Chymosin B is the predominant form.

GENETICS 21/2; 71/3; 111/1; 150/3; 217/2; 255/2; 303/3; 336/3
#introns
CLASSIFICATION #superfamily pepsin
KEYWORDS aspartic proteinase; gastric juice; hydrolase; protein digestion; stomach; zymogen

FEATURE 1-16 #domain signal sequence #status experimental #label STRG\ 17-58 #domain activation peptide #status experimental #label APT\

59-381 #product chymosin #status experimental #label MPT\
92,274 #active_site Asp #status experimental\
103-110,265-269, #disulfide_bonds #status experimental
308-341 #length 381 #molecular-weight 42179 #checksum 5095
SUMMARY

Query Match 45.6%; Score 62; DB 1; Length 381;
Best Local Similarity 64.3%; Pred. No. 5.60e+00;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 88 TVLF-DTGSSDFWV 100
1111:1111
QY 44 TVLFHEPGSSVWV 57

RESULT 12
ENTRY CMSHB #type complete
TITLE chymosin (EC 3.4.23.4) B precursor - sheep
ALTERNATE_NAMES #formal_name Ovis orientalis aries, Ovis ammon aries
ORGANISM #common_name domestic sheep
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-Sep-1997
ACCESSIONS S10996
REFERENCE S10996
#authors Pungercar, J.; Strukelj, B.; Gubensek, F.; Turk, V.; Kregar, I.
#journal Nucleic Acids Res. (1990) 18:4602
#title Complete primary structure of lamb preprochymosin deduced from cDNA.
#cross-references MUID:90356410
#accession S10996
##molecule_type mRNA
##residues 1-381 ##label PUN
##cross-references EMBL:X53037; NID:g1373; PID:g1374

CLASSIFICATION #superfamily pepsin
KEYWORDS aspartic proteinase; gastric juice; hydrolase; protein digestion; stomach

FEATURE 1-16 #domain signal sequence #status predicted #label SIG\
17-58 #domain activation peptide #status predicted #label APT\
59-381 #product chymosin B #status predicted #label MAT\
92,274 #active_site Asp #status predicted\
103-110,265-269, #disulfide_bonds #status predicted
308-341 #length 381 #molecular-weight 42074 #checksum 5159
SUMMARY

Query Match 45.6%; Score 62; DB 1; Length 381;
Best Local Similarity 64.3%; Pred. No. 5.60e+00;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 88 TVLF-DTGSSDFWV 100
1111:1111
QY 44 TVLFHEPGSSVWV 57

RESULT 13
ENTRY A39314 #type complete
TITLE gastricsin (EC 3.4.23.3) precursor - bullfrog
ALTERNATE_NAMES #formal_name Rana catesbeiana #common_name bullfrog
ORGANISM #formal_name Rana catesbeiana #sequence_revision 19-Jun-1992 #text_change 08-Sep-1997
DATE
ACCESSIONS A39314
REFERENCE A39314
#authors Yakabe, E.; Tanji, M.; Ichinose, M.; Goto, S.; Miki, K.; Kurokawa, K.; Ito, H.; Kageyama, T.; Takahashi, K.; J. Biol. Chem. (1991) 266:22436-22443
#journal Purification, characterization, and amino acid sequences of pepsinogens and pepsins from the esophageal mucosa of bullfrog (Rana catesbeiana).
#title
#cross-references MUID:92042186
#accession A39314
##status preliminary
##molecule_type mRNA
##residues 1-384 ##label YAK
#cross-references GB:M73750; NID:g213687; PID:g213688
CLASSIFICATION #superfamily pepsin
KEYWORDS aspartic proteinase; hydrolase; protein digestion
SUMMARY #length 384 #molecular-weight 41764 #checksum 4918

Query Match 45.6%; Score 62; DB 2; Length 384;
Best Local Similarity 61.5%; Pred. No. 5.60e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 82 VLF-DTGSSNLWV 93
1111:1111
QY 45 VLFHEPGSSVWV 57

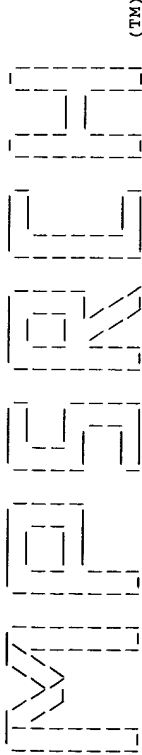
RESULT 14
ENTRY A29937 #type complete
TITLE gastricsin (EC 3.4.23.3) precursor - human
ALTERNATE_NAMES #formal_name Homo sapiens #common_name man
ORGANISM #formal_name Homo sapiens #sequence_revision 17-Oct-1988 #text_change 08-Sep-1997
DATE
ACCESSIONS A29937; A31811; PX0028; I54213; A91125; A23458
REFERENCE A29937
#authors Hayano, T.; Sogawa, K.; Ichihara, Y.; Fujii-Kuriyama, Y.; Takahashi, K.
#journal J. Biol. Chem. (1988) 263:1382-1385
#title Primary structure of human pepsinogen C gene.
#cross-references MUID:88087276
#accession A29937
##molecule_type DNA
##residues 1-388 ##label HAY
REFERENCE A31811
#authors Taggart, R.T.; Cass, L.G.; Mohandas, T.K.; Dorby, P.; Barr,

P.J.; Pals, G.; Bell, G.I.
J. Biol. Chem. (1989) 264:375-379
Human pepsinogen C (progastricsin). Isolation of cDNA clones,
localization to chromosome 6, and sequence homology with
pepsinogen A.
#cross-references MUID:89079679
#accession A31811
#molecule_type mRNA
#residues 1-388 #label TAG
#cross-references GB:J04443; NID:g551175; PID:g551176
REFERENCE PX0023
#authors Athauda, S.B.P.; Tanji, M.; Kageyama, T.; Takahashi, K.
#journal J. Biochem. (1989) 106:920-927
#title A comparative study on the NH2-terminal amino acid sequences
and some other properties of six isozymic forms of human
pepsinogens and pepsins.
#cross-references MUID:90130402
#accession PX0028
#molecule_type protein
#residues 17-101 #label ATH
REFERENCE I54213
#authors Pals, G.; Azuma, T.; Mohandas, T.K.; Bell, G.I.; Bacon, J.;
Samloff, I.M.; Waiz, D.A.; Barr, P.J.; Taggart, R.T.
#journal Genomics (1989) 4:137-148
#title Human pepsinogen C (progastricsin) polymorphism: evidence for
a single locus located at 6p21.1-pter.
#cross-references MUID:89290840
#accession I54213
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-388 #label RES
#cross-references GB:M23077; NID:gi89830; PID:g387015; GB:J03063
#note parts of this sequence, including the amino end and
carboxyl ends of the mature protein, were determined
by protein sequencing
REFERENCE A91125
#authors Foltmann, B.; Jensen, A.L.
#journal Eur. J. Biochem. (1982) 128:63-70
#cross-references MUID:83079318
#accession A91125
#molecule_type protein
#residues 17-39, 'ED', 42-51, 'S', 53-64 #label FOL
#note pro-form; 29-Leu was also found
#note activation at pH 2 is proposed to involve conformation
change, cleavage after Phe-42, and cleavage after
Leu-59
GENETICS
#gene GDB:PGC
#cross-references GDB:119485; OMIM:169740
#map_position 6p21.3-6p21.1
#introns 20/2; 70/3; 110/1; 149/3; 216/2; 256/2; 305/3; 338/3
CLASSIFICATION #superfamily pepsin
KEYWORDS aspartic proteinase; hydrolase; protein digestion; stomach;
zymogen
FEATURE
1-16 #domain signal sequence #status predicted #label SIG\
17-59 #domain propeptide #status experimental #label PRO\
60-388 #product gastricsin #status experimental #label MAT
SUMMARY
#length 388 #molecular-weight 42426 #checksum 9699
Query Match 45.6%; Score 62; DB 2; Length 388;
Best Local Similarity 61.5%; Pred. No. 5.60e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;
Db 88 VLF-DTGSNLMW 99
||| : |||: |||
QY 45 VLFHEPGSSSVWV 57
RESULT 15
ENTRY JE0371 #type complete
TITLE pepsinogen C - Chicken
ORGANISM #formal_name Gallus gallus #common_name chicken

21-Jan-1999 #sequence_revision 21-Jan-1999 #text_change
21-Jan-1999
JE0371
JE0370
Sakamoto, N.; Saiga, H.; Yasugi, S.
Biochem. Biophys. Res. Commun. (1998) 250:420-424
Analysis of temporal expression pattern and cis-regulatory
sequences of chicken pepsinogen A and C.
JE0371
#accession preliminary
#status 1-389 #label SAK
#residues
SUMMARY
#length 389 #molecular-weight 42510 #checksum 9303
Query Match 45.6%; Score 62; DB 3; Length 389;
Best Local Similarity 61.5%; Pred. No. 5.60e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;
Db 88 VLF-DTGSNLMW 99
||| : |||: |||
QY 45 VLFHEPGSSSVWV 57

Search completed: Thu Jul 8 18:11:24 1999
Job time : 9 secs.

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:10:20 1999; MasPar time 3.51 Seconds
Tabular output not generated. 145.172 Million cell updates/sec

Title: >US-09-041-236-2
(42-59) from US09041236.pep (4 of 45)
Perfect Score: 136
Sequence: 1 PHTVLFHEPGSSSWVGG 18

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 28.598; Variance 41.609; scale 0.687

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	65	47.8	PERT_PIG	THYROID PEROXIDASE PRE	7.16e+01
2	64	47.1	PEPC_RAT	PROGASTRINSIN PRECURSO	1.08e+00
3	63	46.3	ULB7_HCMVA	HYPOTHETICAL PROTEIN U	1.61e+00
4	62	45.6	PEPC_WACFU	PROGASTRINSIN PRECURSO	2.41e+00
5	62	45.6	CHYM_BOVIN	PROCHYMOSIN A/B PRECUR	2.41e+00
6	62	45.6	CHYM_SHEEP	PROCHYMOSIN PRECURSOR	2.41e+00
7	62	45.6	PEPC_HUMAN	PROGASTRINSIN PRECURSO	2.41e+00
8	62	45.6	PEPC_CAVPO	PROGASTRINSIN PRECURSO	2.41e+00
9	61	44.9	PEPA_BOVIN	PEPSINOGEN A (EC 3.4.2	3.57e+00
10	61	44.9	CDSA_ECOLI	PHOSPHATIDATE CYTIDILY	3.57e+00
11	61	44.9	PEPA_PIG	PEPSINOGEN A PRECURSOR	3.57e+00
12	61	44.9	PEPA_RABIT	PEPSINOGEN II-4 PRECUR	3.57e+00
13	61	44.9	PEP2_RABIT	PEPSINOGEN II-2/3 PREC	3.57e+00
14	61	44.9	PEP1_RABIT	PEPSINOGEN II-1 PRECUR	3.57e+00
15	61	44.9	PEP3_RABIT	PEPSINOGEN III PRECURS	3.57e+00
16	61	44.9	PEPA_MACMU	PEPSINOGEN A-2 OR A-3	3.57e+00
17	61	44.9	PEP2_WACFU	PEPSINOGEN A-1 PRECURS	3.57e+00
18	61	44.9	PEP1_WACFU	CATHEPSIN E PRECURSOR	3.57e+00
19	61	44.9	CATE_CAVPO	CATHEPSIN E PRECURSOR	3.57e+00
20	61	44.9	CATE_HUMAN	CATHEPSIN E PRECURSOR	3.57e+00
21	61	44.9	CATE_MOUSE	CATHEPSIN E PRECURSOR	3.57e+00
22	61	44.9	CATE_RAT	CATHEPSIN E PRECURSOR	3.57e+00
23	61	44.9	REN1_HUMAN	RENIN PRECURSOR, RENAL	3.57e+00

24	61	44.9	1	CYP1_CVNCA	CVPROSIN PRECURSOR (EC	3.57e+00
25	61	44.9	496	ASPR_ORYSA	ASPARTIC PROTEINASE PR	3.57e+00
26	61	44.9	508	ASPR_HORVU	PHYTESIN PRECURSOR (E	3.57e+00
27	61	44.9	509	APRL_ORYSA	ASPARTIC PROTEINASE OR	3.57e+00
28	61	44.9	513	ASPR_CUCPE	ASPARTIC PROTEINASE PR	3.57e+00
29	61	44.9	1124	YPHG_ECOLI	HYPOTHETICAL 127.3 KD	3.57e+00
30	61	44.9	2334	WAPA_BACSU	WALL-ASSOCIATED PROTEI	3.57e+00
31	60	44.1	345	CATD_PIG	CATHEPSIN D (EC 3.4.23	5.28e+00
32	60	44.1	383	PEPE_CHICK	EMBRIONIC PEPSINOGEN P	5.28e+00
33	60	44.1	388	PEPA_HUMAN	PEPSINOGEN A PRECURS	5.28e+00
34	60	44.1	388	PEP4_MACFU	PEPSINOGEN A-4 PRECURS	5.28e+00
35	60	44.1	398	CATD_CHICK	CATHEPSIN D PRECURSOR	5.28e+00
36	60	44.1	407	CATD_RAT	CATHEPSIN D PRECURSOR	5.28e+00
37	60	44.1	410	CATD_MOUSE	CATHEPSIN D PRECURSOR	5.28e+00
38	60	44.1	412	CATD_HUMAN	CATHEPSIN D PRECURSOR	5.28e+00
39	60	44.1	587	BARI_YEAST	BARRIERPEPSIN PRECURSO	5.28e+00
40	60	44.1	712	PERL_BOVIN	LACTOPEROXIDASE PRECUR	5.28e+00
41	60	44.1	933	PERT_HUMAN	THYROID PEROXIDASE PRE	5.28e+00
42	59	43.4	508	YL12_YEAST	PUTATIVE ASPARTYL PROT	7.77e+00
43	58	42.6	303	Y247_HUMAN	HYPOTHETICAL PROTEIN K	1.14e+01
44	58	42.6	596	MKC7_YEAST	ASPARTIC PROTEINASE MK	1.14e+01
45	58	42.6	761	YL31_CAEEL	HYPOTHETICAL 86.0 KD P	1.14e+01

ALIGNMENTS

RESULT 1	PERT_PIG	STANDARD;	PRT;	926 AA.
AC	P09933;			
DT	01-MAR-1989 (REL. 10, CREATED)			
DT	01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)			
DT	01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)			
DE	THYROID PEROXIDASE PRECURSOR (EC 1.11.1.8) (TPO).			
GN	TPO.			
OS	SUS SCROFA (PIG).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC	ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 88007634.			
RA	MAGNUSSON R.P., GESTAUTAS J., TAUROG A., RAPOPORT B.;			
RT	"Molecular cloning of the structural gene for porcine thyroid peroxidase."			
RL	J. BIOL. CHEM. 262:13885-13888(1987).			
RN	[2]			
RP	SEQUENCE OF 595-926 FROM N.A.			
RX	MEDLINE; 87054611.			
RA	MAGNUSSON R.P., GESTAUTAS J., SETO P., TAUROG A., RAPOPORT B.;			
RT	"Isolation and characterization of a cDNA clone for porcine thyroid peroxidase."			
RL	FEB5 LETT. 208:391-396(1986).			
CC	-1- FUNCTION: IODINATION AND COUPLING OF THE HORMONOGENIC TYROSINES IN THYROGLOBULIN TO YIELD THE THYROID HORMONES T(3) AND T(4).			
CC	-1- CATALYTIC ACTIVITY: IODIDE + H(2)O(2) -> IODINE + 2 H(2)O.			
CC	-1- COFACTOR: HEME (PROTOPHYRIN IX).			
CC	-1- PATHWAY: THYROID HORMONE BIOSYNTHESIS.			
CC	-1- SUBCELLULAR LOCATION: MEMBRANE.			
CC	-1- SIMILARITY: HIGH WITH EOSINOPHIL PEROXIDASE (EPO), AND MYELOPEROXIDASE (MPO).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; X04645; G2142;			
CC	PIR; A27416; OPPGIT.			
DR	PROSITE; PS00436; PEROXIDASE.2; FALSE_NEG.			
DR	PROSITE; PS00435; PEROXIDASE_1; 1.			
DR	PFAM; PF00008; EGF; 1.			

```
DR PFAM; PF00084; sushi; 1.
DR PFAM; PF00141; peroxidase; 1.
DR HSP; P05164; IMHL.
KW OXIDOREDUCTASE; PEROXIDASE; HEME; TRANSMEMBRANE; GLYCOPROTEIN;
KW SIGNAL.
FT SIGNAL. 1 14 POTENTIAL.
FT CHAIN 15 926 THYROID PEROXIDASE.
FT ACT_SITE 239 239 DISTAL HISTIDINE (POTENTIAL).
FT ACT_SITE 395 395 DISTAL ARGININE (POTENTIAL).
FT ACT_SITE 493 493 PROXIMAL HISTIDINE (HEME AXIAL LIGAND)
(POTENTIAL).
FT TRANSEM 845 869 POTENTIAL.
FT CARBOHYD 129 129 POTENTIAL.
FT CARBOHYD 277 277 POTENTIAL.
FT CARBOHYD 307 307 POTENTIAL.
FT CARBOHYD 342 342 POTENTIAL.
SQ SEQUENCE 926 AA; 100442 MW; 19C971EC CRC32;

Query Match 47.8%; Score 65; DB 1; Length 926;
Best Local Similarity 42.9%; Pred. No. 7.16e-01;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 623 LYOHDPNDVWLG 636
Qy 46 LFHPGSSVWVG 59
I: | | | | |
: | | | | |

RESULT 2
ID PEPC.RAT STANDARD; PRT; 392 AA.
AC P04073;
DT 01-NOV-1986 (REL. 03, CREATED)
DT 01-NOV-1986 (REL. 03, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROGASTRICIN PRECURSOR (EC 3.4.23.3) (PEPSINOGEN C).
GN PGC.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN WISTAR.
RX MEDLINE; 89255508.
RA ISHIGAKI T., ICHIHARA Y., HAYANO T., KATSURA I., SOGAWA K.,
RA FUJII-KURIYAMA Y., TAKAHASHI K.;
RT "Primary structure and transcriptional regulation of rat pepsinogen C
gene.";
RL J. BIOL. CHEM. 264:10193-10199(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN WISTAR.
RX MEDLINE; 87054020.
RA ICHIHARA Y., SOGAWA K., MOROHASHI K., FUJII-KURIYAMA Y., TAKAHASHI K.;
RT "Nucleotide sequence of a nearly full-length cDNA coding for
pepsinogen of rat gastric mucosa.";
RL EUR. J. BIOCHEM. 161:7-12(1986).
RN [3]
RP SEQUENCE OF 16-112.
RC STRAIN WISTAR.
RX MEDLINE; 84257697.
RA ARAI K.M., MUTO N., TANI S., AKAHANE K.;
RT "The N-terminal sequence of rat pepsinogen.";
RL BIOCHIM. BIOPHYS. ACTA 788:256-261(1984).
CC -1- CATALYTIC ACTIVITY: MORE RESTRICTED SPECIFICITY THAN PEPSIN A, BUT
SHOWS PREFERENTIAL CLEAVAGE AT TYR-|-XAA BONDS; HIGH ACTIVITY
TOWARDS HEMOGLOBIN AS SUBSTRATE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
EUKARYOTIC ASPARTYL PROTEASES FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; M25993; G206083;
CC EMBL; M25985; G206083; JOINED.
CC EMBL; M25986; G206083; JOINED.
CC EMBL; M25987; G206083; JOINED.
CC EMBL; M25988; G206083; JOINED.
CC EMBL; M25989; G206083; JOINED.
CC EMBL; M25990; G206083; JOINED.
CC EMBL; M25991; G206083; JOINED.
CC EMBL; M25992; G206083; JOINED.
CC EMBL; X04644; G56881;
CC PIR; A24608; A24608.
CC PIR; A05145; A05145.
CC PIR; A33510; A33510.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
CC PFAM; PF00026; asp; 1.
CC HSP; P20142; IAVE.
KW HYDROLASE; ASPARTYL PROTEASE; ZMOGEN; DIGESTION; STOMACH;
KW GASTRIC JUICE; SIGNAL.
FT SIGNAL 1 16 ACTIVATION PEPTIDE.
FT PROPEP 17 62 GASTRICIN C.
FT CHAIN 63 392
FT ACT_SITE 94 94
FT ACT_SITE 280 280
FT DISULFID 107 112 BY SIMILARITY.
FT DISULFID 270 275 BY SIMILARITY.
FT DISULFID 314 347 BY SIMILARITY.
FT CONFLICT 31 31 E -> Q (IN REF. 3).
FT CONFLICT 103 103 S -> A (IN REF. 3).
FT CONFLICT 109 109 S -> L (IN REF. 3).
SQ SEQUENCE 392 AA; 42833 MW; 80F80F62 CRC32;

Query Match 47.1%; Score 64; DB 1; Length 392;
Best Local Similarity 53.3%; Pred. No. 1.08e+00;
Matches 8; Conservative 5; Mismatches 1; Indels 1; Gaps 1;

Db 91 VLF-DTGSSNLWVSS 104
Qy 45 VLFHPGSSVWVG 59
I: | | | | |
: | | | | |

RESULT 3
ID ULB7_HCMVA STANDARD; PRT; 424 AA.
AC P16770;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN ULL117.
GN ULL117.
OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE.
OC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90269039.
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTINETTI J.A.,
RA PREDDIE E., SATCHEL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT "Analysis of the protein-coding content of the sequence of human
cytomegalovirus strain AD169.";
RL CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-6 AND
HSV-7 U84 AND HCMV ULL117.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL: X17403; G59727; -
DR PIR: S09884; S09884;
KW HYPOTHETICAL PROTEIN.
FT CARBOHYD 247 247 POTENTIAL.
FT CARBOHYD 309 309 POTENTIAL.
FT CARBOHYD 396 396 POTENTIAL.
SQ SEQUENCE 424 AA; 45463 MW; 3412F873 CRC32;

Query Match 46.3%; Score 63; DB 1; Length 424;
Best Local Similarity 38.9%; Pred. No. 1.61e+00;
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Db 352 PRSTSYHETGVYQMWVG 369
QY 42 PHTVLFHEPGSSVWVG 59

RESULT 4
ID PEPC_MAFU STANDARD; PRT; 377 AA.
AC P03955;
DT 23-OCT-1986 (REL. 02, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROGASTRICIN PRECURSOR (EC 3.4.23.3) (PEPSINOGEN C) (FRAGMENT).
GN PCG.
OS MACACA FUSCATA FUSCATA (JAPANESE MACAQUE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=GASTRIC MUCOSA;
RX MEDLINE; 92037645
RA KAGEYAMA T., TANABE K., KOINAI O.;
RT "Development-dependent expression of isozymogens of monkey
RT pepsinogens and structural differences between them.";
RL EUR. J. BIOCHEM. 202:205-215(1991).
[2]
RN SEQUENCE OF 6-377.
RX MEDLINE; 86168133.
RA KAGEYAMA T., TAKAHASHI K.;
RT "The complete amino acid sequence of monkey progastricsin.";
RL J. BIOL. CHEM. 261:4406-4419(1986).
[3]
RN SEQUENCE OF 6-65.
RX MEDLINE; 85289106.
RA KAGEYAMA T., TAKAHASHI K.;
RT "Monkey pepsinogens and pepsins. VII. Analysis of the activation
RT process and determination of the NH2-terminal 60-residue sequence of
RT Japanese monkey progastricsin, and molecular evolution of
RT pepsinogens.";
RL J. BIOCHEM. 97:1235-1246(1985).
CC -1- CATALYTIC ACTIVITY: MORE RESTRICTED SPECIFICITY THAN PEPSIN A, BUT
CC SHOWS PREFERENTIAL CLEAVAGE AT TYR-I-XAA BONDS; HIGH ACTIVITY
CC TOWARDS HEMOGLOBIN AS SUBSTRATE.
CC -1- THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY HORMONES AND
CC RELATED SUBSTANCES.
CC -1- PTM: EACH PEPSINOGEN IS CONVERTED TO CORRESPONDING PEPSIN AT PH
CC 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA ACTIVATION
CC SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC CLEAVAGE
CC VIA AN INTERMEDIATE FORM(S).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X59754; G38073; -
DR -----
```

```
DR PIR: A00986; PEMOCJ.
DR PIR: A22402; A22402.
DR PIR: S19683; S19683.
DR PIR: S16066; S16066.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR PFAM; PF00026; asp; 1.
DR HSP; P20142; LAVF.
KW HYDROLASE; ASPARTYL PROTEASE; ZYMOGEN; DIGESTION; STOMACH;
KW GASTRIC JUICE; SIGNAL.
FT NON_TER 1 1
FT SIGNAL <1 5
FT PROPEP 6 31 ACTIVATION PEPTIDE.
FT PROPEP 32 48 ACTIVATION PEPTIDE.
FT CHAIN 49 377 GASTRICIN.
FT DISULFID 93 98
FT DISULFID 256 260
FT DISULFID 299 332
FT ACT_SITE 80 80
FT ACT_SITE 265 265
FT ACT_SITE 331 331 Y -> V (IN REF. 2).
FT CONFLICT 349 349 L -> LVY (IN REF. 2).
SQ SEQUENCE 377 AA; 41148 MW; 199544EE CRC32;

Query Match 45.6%; Score 62; DB 1; Length 377;
Best Local Similarity 61.5%; Pred. No. 2.41e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 77 VLF-DTGSSNLWV 88
QY 45 VLFHEPGSSVWV 57

RESULT 5
ID CHYM_BOVIN STANDARD; PRT; 381 AA.
AC P00794;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PROCHYMOSIN A/B PRECURSOR (EC 3.4.23.4) (PREPRORENNIN).
GN CYM OR CPC.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
[1]
RN SEQUENCE FROM N.A. (CHYMOSIN B).
RX MEDLINE; 86301873.
RA HIDAKA M., SASAKI K., UOZUMI T., BEPPU T.;
RT "Cloning and structural analysis of the calf prochymosin gene.";
RL GENE 43:197-203(1986).
[2]
RN SEQUENCE FROM N.A. (CHYMOSIN B).
RX MEDLINE; 82221400.
RA HARRIS T.J.R., LOWE P.A., LYONS A., THOMAS P.G., EATON M.A.W.,
RA MILLICAN T.A., PATEL T.P., BOSE C.C., CAREY N.H., DOEL M.T.;
RT "Molecular cloning and nucleotide sequence of cDNA coding for calf
RT preprochymosin.";
RL NUCLEIC ACIDS RES. 10:2177-2187(1982).
[3]
RN SEQUENCE OF 59-381 (CHYMOSIN B), AND DISULFIDE BONDS.
RX MEDLINE; 79239460.
RA FOLTMANN B., PEDERSEN V.B., KAUFFMAN D., WYBRANDT G.;
RT "The primary structure of calf chymosin.";
RL J. BIOL. CHEM. 254:8447-8456(1979).
[4]
RN SEQUENCE OF 17-77 (CHYMOSIN B).
RX MEDLINE; 76022411.
RA PEDERSEN V.B., FOLTMANN B.;
RT "Amino-acid sequence of the peptide segment liberated during
RT activation of prochymosin (prorennin).";
RL EUR. J. BIOCHEM. 55:95-103(1975).
[5]
RN CHYMOSIN B, ACTIVE SITE PEPTIDES.
RX MEDLINE; 75060332.
DR -----
```

RA CHANG W.-J., TAKAHASHI K.;
 RT "The structure and function of acid proteases. III. Isolation and
 RT characterization of the active-site peptides from bovine rennin.";
 RL J. BIOCHEM. 76:467-474(1974).
 RN [6]
 RP SEQUENCE FROM N.A. (CHYMOSIN A).
 RX MEDLINE: 83054629.
 RA MOIR D., MAO J., SCHUMM J.W., VOVIS G.F., ALFORD B.L.,
 RA TAUNTON-RIGBY A.;
 RT "Molecular cloning and characterization of double-stranded cDNA
 RT coding for bovine chymosin.";
 RL GENE 19:127-138(1982).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE: 92046065.
 RA NEWMAN M., SAFRO M., FRAZAO C., KAHN G., ZDANOV A., TICKLE I.J.,
 RA BLUNDELL T.L., ANDREVA N.;
 RT "X-ray analyses of aspartic proteinases. IV. Structure and refinement
 RT at 2.2-A resolution of bovine chymosin.";
 RL J. MOL. BIOL. 221:1295-1309(1991).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF MUTANT.
 RX MEDLINE: 91104895.
 RA STROP P., SEDLACEK J., STYS J., KADERABKOVA Z., BLAHA I.,
 RA PAVLICKOVA L., POHL J., FABRY M., KOSTKA V., NEWMAN M., FRAZAO C.,
 RA SHEARER A., TICKLE I.J., BLUNDELL T.L.;
 RT "Engineering enzyme subsite specificity: preparation, kinetic
 RT characterization, and X-ray analysis at 2.0-A resolution of Val111Phe
 RT site-mutated calf chymosin";
 RL BIOCHEMISTRY 29:9863-9871(1990).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RX MEDLINE: 91017501.
 RA GILLILAND G.L., WINBORNE E.L., NACHMAN J., WLODAWER A.;
 RT "The three-dimensional structure of recombinant bovine chymosin at
 RT 2.3-A resolution.";
 RL PROTEINS 8:82-101(1990).
 RN [10]
 RP MUTAGENESIS OF CYS-308 AND CYS-341.
 RX MEDLINE: 92412108.
 RA HUANG K., ZHANG Z., LIU N., ZHANG Y., ZHANG G., YANG K.;
 RT "Functional implication of disulfide bond, Cys250-Cys283, in bovine
 RT chymosin";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 187:692-696(1992).
 CC -!- FUNCTION: CHYMOSIN IS SYNTHESIZED IN THE MUCOSA OF THE ABOMASUM
 CC (FOURTH STOMACH) OF YOUNG (UNWEANED) RUMINANTS. THE ENZYME
 CC HYDROLYSES CASEIN TO PARACASEIN.
 CC -!- CATALYTIC ACTIVITY: BROAD SPECIFICITY SIMILAR TO THAT OF PEP SIN A.
 CC CLOTS MILK BY CLEAVAGE OF A SINGLE BOND IN CASEIN (KAPPA CHAIN).
 CC -!- SUBUNIT: MONOMER.
 CC -!- POLYMORPHISM: FORMS A AND B ARE PROBABLY ALLELIC VARIANTS. FORM B
 CC IS THE PREDOMINANT FORM AND IS THE SEQUENCE SHOWN.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14077; G162856; -;
 DR EMBL: M14069; G162856; JOINED.
 DR EMBL: M14070; G162856; JOINED.
 DR EMBL: M14071; G162856; JOINED.
 DR EMBL: M14072; G162856; JOINED.
 DR EMBL: M14073; G162856; JOINED.
 DR EMBL: M14074; G162856; JOINED.
 DR EMBL: M14075; G162856; JOINED.
 DR EMBL: J00002; G162858; -;
 DR EMBL: J00003; G162860; -;
 DR EMBL: J00003; E18172; ALT_SEQ.
 DR EMBL: J00003; E18173; ALT_INIT.
 DR PIR: A00985; CMBO.
 DR PIR: A25631; A25631.
 DR PDB: 1CMS; 15-JUL-92.
 DR PDB: 3CMS; 15-OCT-92.
 DR PDB: 4CMS; 15-OCT-91.
 DR PDB: 1CZ1; 01-APR-97.
 DR PROSITE: PS00141; ASP_PROTEASE; 2.
 DR PFAM: PF00026; asp; 1
 KW HYDROLASE; ASPARTYL PROTEASE; DIGESTION; STOMACH; GASTRIC JUICE;
 KW ZIMOGEN; SIGNAL; 3D-STRUCTURE.
 FT SIGNAL 1
 FT PROPEP 17 58 ACTIVATION PEPTIDE.
 FT CHAIN 59 381 CHYMOSIN B.
 FT ACT_SITE 92 92
 FT ACT_SITE 274 274
 FT REPEAT 92 102
 FT REPEAT 274 284
 FT DISULFID 105 110
 FT DISULFID 265 269
 FT DISULFID 308 341
 FT VARIANT 302 302
 FT CONFLICT 17 17
 FT CONFLICT 109 109
 FT CONFLICT 155 155
 FT CONFLICT 218 218
 FT CONFLICT 230 230
 FT CONFLICT 232 232
 FT CONFLICT 302 302
 FT STRAND 62 69
 FT TURN 70 72
 FT STRAND 73 80
 FT TURN 81 84
 FT STRAND 85 92
 FT TURN 93 94
 FT STRAND 98 102
 FT TURN 103 104
 FT TURN 112 112
 FT TURN 112 112
 FT STRAND 116 116
 FT TURN 118 120
 FT STRAND 122 123
 FT STRAND 125 135
 FT TURN 136 137
 FT STRAND 138 151
 FT TURN 152 153
 FT STRAND 154 166
 FT TURN 170 174
 FT STRAND 179 182
 FT TURN 186 188
 FT TURN 191 192
 FT HELIX 196 202
 FT TURN 203 204
 FT STRAND 210 214
 FT TURN 217 218
 FT STRAND 222 226
 FT HELIX 231 233
 FT STRAND 234 242
 FT STRAND 246 246
 FT TURN 247 248
 FT STRAND 249 258
 FT TURN 259 260
 FT STRAND 261 264
 FT TURN 266 267
 FT STRAND 269 273
 FT TURN 275 276
 FT STRAND 280 282
 FT HELIX 284 294
 FT TURN 295 295
 FT STRAND 297 299
 FT TURN 300 302
 FT STRAND 303 306
 G -> D (IN CHYMOSIN A).
 T -> A (IN REF. 2 AND 6).
 A -> G (IN REF. 1).
 V -> L (IN REF. 6).
 N -> D (IN REF. 3).
 D -> N (IN REF. 2).
 S -> C (IN REF. 6).
 G -> D (IN REF. 6).

FT TURN 308 309
FT TURN 311 313
FT STRAND 317 321
FT TURN 322 323
FT STRAND 324 328
FT HELIX 330 333
FT STRAND 334 337
FT TURN 338 339
FT STRAND 340 343
FT STRAND 345 347
FT STRAND 354 356
FT HELIX 358 361
FT TURN 362 363
FT STRAND 364 369

Note: remainder of annotations omitted.

Query Match 45.6%; Score 62; DB 1; Length 381;
Best Local Similarity 64.3%; Pred. No. 2.41e+00;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 88 TVLF-DTGSDFWV 100
||||: ||| ||
QY 44 TVLFHEPGSSVWV 57

RESULT 6 STANDARD; PRT: 381 AA.

AC P18276;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROCHYMOSIN PRECURSOR, B-FORM (EC 3.4.23.4) (PREPRORENNIN).
GN CYM.
OS OVIS ARIES (SHEEP).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; CAPRINAE; OVIS.
RN [1]
RP SEQUENCE FROM N.A.
RX PUNGECAR J., STRURELJ B., GUBENSEK F., TURK V., KREGAR I.;
RT "Complete primary structure of lamb preprochymosin deduced from
CDNA.";
RL NUCLEIC ACIDS RES. 18:4602-4602(1990).
CC -!- FUNCTION: CHYMOSIN IS SYNTHESIZED IN THE MUCOSA OF THE STOMACH.
CC THE ENZYME HYDROLYZES CASEIN TO PARACASEIN.
CC -!- CATALYTIC ACTIVITY: BROAD SPECIFICITY SIMILAR TO THAT OF PEPSIN A.
CC CLOTS MILK BY CLEAVAGE OF A SINGLE BOND IN CASEIN (KAPPA CHAIN).
CC -!- SUBUNIT: MONOMER.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X53037; GI374; -
CC PIR: S10996; CMSHB.
CC PROSITE: PS00141; ASP_PROTEASE; 2.
CC PFAM: PF00026; asp; 1.
CC HSSP: P00794; 1CMS.

KW HYDROLASE; ASPARTYL PROTEASE; DIGESTION; STOMACH; GASTRIC JUICE;
KW ZMOGEN; SIGNAL.
FT SIGNAL 1 16
FT PROPEP 17 58 ACTIVATION PEPTIDE.
FT CHAIN 59 381 CHYMOSIN.
FT ACT_SITE 92 92 BY SIMILARITY.
FT ACT_SITE 274 274 BY SIMILARITY.
FT REPEAT 92 102

FT REPEAT 274 284
FT DISULFID 105 110 BY SIMILARITY.
FT DISULFID 265 269 BY SIMILARITY.
FT DISULFID 308 341 BY SIMILARITY.
SQ SEQUENCE 381 AA; 42074 MW; 92A0AB58 CRC32;

Query Match 45.6%; Score 62; DB 1; Length 381;
Best Local Similarity 64.3%; Pred. No. 2.41e+00;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Db 88 TVLF-DTGSDFWV 100
||||: ||| ||
QY 44 TVLFHEPGSSVWV 57

RESULT 7 STANDARD; PRT: 388 AA.

AC P20142;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROGASTRICIN PRECURSOR (EC 3.4.23.3) (PEPSINOGEN C).
GN PGC.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88087276.
RA HAYANO T., SOGAWA K., ICHIHARA Y., FUJII-KURIYAMA Y., TAKAHASHI K.;
RT "Primary structure of human pepsinogen C gene."
RL J. BIOL. CHEM. 263:1382-1385(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89079679.
RA TAGGART R.T., CASS L.G., MOHANDAS T.K., DERBY P., BARR P.J., PALS G.,
RA BELL G.I.;
RT "Human pepsinogen C (progastricisin). Isolation of cDNA clones,
RT localization to chromosome 6, and sequence homology with pepsinogen
RT A.";
RL J. BIOL. CHEM. 264:375-379(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE; 89290840.
RA PALS G., AZUMA T., MOHANDAS T.K., BELL G.I., BACON J.,
RA SAMLOFF I.M., WALZ D.A., BARR P.J., TAGGART R.T.;
RT "Human pepsinogen C (progastricisin) polymorphism: evidence for a
RT single locus located at 6p21.1-pter."
RL GENOMICS 4:137-148(1989).
RN [4]
RP SEQUENCE FROM N.A.
RA WONG R.N.S., TANG J.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP SEQUENCE OF 17-101.
RX MEDLINE; 90130402.
RA ATHAUDA S.B.P., TANJI M., KAGEYAMA T., TAKAHASHI K.;
RT "A comparative study on the NH2-terminal amino acid sequences and
RT some other properties of six isozymic forms of human pepsinogens and
RT pepsins.";
RL J. BIOCHEM. 106:920-927(1989).
RN [6]
RP SEQUENCE OF 17-64.
RX MEDLINE; 83079318.
RA FOLTMANN B., JENSEN A.L.;
RT "Human progastricisin. Analysis of intermediates during activation
RT into gastricsin and determination of the amino acid sequence of the
RT propeptide.";
RL EUR. J. BIOCHEM. 128:63-70(1982).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.62 ANGSTROMS).
RX MEDLINE; 95230687.

RA MOORE S.A., SIELECKI A.R., CHERNAIA M.M., TARASOVA N.I., JAMES M.N.G.;
RT "Crystal and molecular structures of human progastricsin at 1.62-A
RL resolution."; J. MOL. BIOL. 247:466-485(1995).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.36 ANGSTROMS).
RX MEDLINE; 98069649.
RA KHAN A.R., CHERNEY M.M., TARASOVA N.I., JAMES M.N.;
RT "Structural characterization of activation 'intermediate 2' on the
RL pathway to human gastricsin."; NAT. STRUCT. BIOL. 4:1010-1015(1997).
RN NAT. STRUCT. BIOL. 4:1010-1015(1997).
CC -!- CATALYTIC ACTIVITY: MORE RESTRICTED SPECIFICITY THAN PERSIN A, BUT
CC SHOWS PREFERENTIAL CLEAVAGE AT TYR-|-XAA BONDS; HIGH ACTIVITY
CC TOWARDS HEMOGLOBIN AS SUBSTRATE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M18667; G387014; ALT_INIT.
DR EMBL; M18659; G387014; JOINED.
DR EMBL; M18660; G387014; JOINED.
DR EMBL; M18661; G387014; JOINED.
DR EMBL; M18662; G387014; JOINED.
DR EMBL; M18663; G387014; JOINED.
DR EMBL; M18664; G387014; JOINED.
DR EMBL; M18665; G387014; JOINED.
DR EMBL; M18666; G387014; JOINED.
DR EMBL; M23077; G387015; -.
DR EMBL; M23069; G387015; JOINED.
DR EMBL; M23070; G387015; JOINED.
DR EMBL; M23071; G387015; JOINED.
DR EMBL; M23072; G387015; JOINED.
DR EMBL; M23073; G387015; JOINED.
DR EMBL; M23074; G387015; JOINED.
DR EMBL; M23075; G387015; JOINED.
DR EMBL; J04443; G531176; -.
DR EMBL; U75272; G1658286; -.
DR PIR; A23458; A23458.
DR PIR; A29937; A29937.
DR PIR; A31811; A31811.
DR PIR; PX0028; PX0028.
DR PDB; 1HTR; 26-JAN-95.
DR PDB; 1AVF; 25-FEB-98.
DR MIM; 169740; -.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR PFAM; PF00026; asp; 1.
KW HYDROLASE; ASPARTYL PROTEASE; ZIMOGEN; DIGESTION; STOMACH;
KW GASTRIC JUICE; SIGNAL; 3D-STRUCTURE.
FT SIGNAL 1 16
FT PROPEP 17 59 ACTIVATION PEPTIDE.
FT CHAIN 60 388 GASTRICIN.
FT ACT_SITE 91 91
FT ACT_SITE 276 276
FT DISULFID 104 109
FT DISULFID 267 271
FT DISULFID 310 343
FT CONFLICT 40 41 GE -> ED (IN REF. 6).
FT CONFLICT 52 52 W -> S (IN REF. 6).
SQ SEQUENCE 388 AA; 42426 MW; F5509B29 CRC32;

Query Match 45.6%; Score 62; DB 1; Length 388;
Best Local Similarity 61.5%; Pred. No. 2.41e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 88 VLF-DTGSSNLWV 99
||| : |||:||

QY 45 VLFHEPGSSVWV 57

RESULT 8
ID PEPC_CAVPO STANDARD; PRT; 394 AA.
AC Q64411;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROGASTRICSIN PRECURSOR (EC 3.4.23.3) (PEPSINOGEN C).
GN PGK.
OS CAVIA PORCELLUS (GUINEA PIG).
OC EURKYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; HYSTRICOGNATHI; CAVIIDAE; CAVIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92355614.
RA KAGEYAMA T., ICHINOSE M., TSUKADA S., MIKI K., KUROKAWA K., KOIWA O.,
RA TANJI M., YABAE E., ATHAUDA S.B., TAKAHASHI K.;
RT "Gastric procathepsin E and progastricsin from guinea pig.
RT Purification, molecular cloning of cDNAs, and characterization of
RT enzymatic properties, with special reference to procathepsin E."; J.
RL J. BIOL. CHEM. 267:16450-16459(1992).
CC -!- CATALYTIC ACTIVITY: MORE RESTRICTED SPECIFICITY THAN PERSIN A, BUT
CC SHOWS PREFERENTIAL CLEAVAGE AT TYR-|-XAA BONDS; HIGH ACTIVITY
CC TOWARDS HEMOGLOBIN AS SUBSTRATE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M88652; G191297; -.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR PFAM; PF00026; asp; 1.
DR HSSP; P20142; 1AVF.
KW HYDROLASE; ASPARTYL PROTEASE; ZIMOGEN; DIGESTION; STOMACH;
KW GASTRIC JUICE; SIGNAL.
FT SIGNAL 1 16
FT PROPEP 17 65 ACTIVATION PEPTIDE.
FT CHAIN 66 394 GASTRICIN C.
FT ACT_SITE 97 97 BY SIMILARITY.
FT ACT_SITE 283 283 BY SIMILARITY.
FT DISULFID 110 115 BY SIMILARITY.
FT DISULFID 273 277 BY SIMILARITY.
FT DISULFID 316 349 BY SIMILARITY.
SQ SEQUENCE 394 AA; 42995 MW; EC48CAE6 CRC32;

Query Match 45.6%; Score 62; DB 1; Length 394;
Best Local Similarity 61.5%; Pred. No. 2.41e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 94 VLF-DTGSSNLWV 105
||| : |||:||
QY 45 VLFHEPGSSVWV 57

RESULT 9
ID PEPA_BOVIN STANDARD; PRT; 110 AA.
AC P00792;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PEPSINOGEN A (EC 3.4.23.1) (FRAGMENT).
GN PGA.
OS BOS TAURUS (BOVINE).
OC EURKYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.


```

RN  [1]
RP  SEQUENCE OF 1-47.
RX  MEDLINE; 74287628.
RA  HARBOE M.K., ANDERSEN P.M., FOLTMANN B., KAY J., KASSELL B.:
RT  "The activation of bovine pepsinogen. Sequence of the peptides
RL  released, identification of a pepsin inhibitor."
RN  J. BIOL. CHEM. 249:4487-4494(1974).
[2]
RP  SEQUENCE OF 41-110.
RX  MEDLINE; 76210768.
RA  HARBOE M.K., FOLTMANN B.:
RT  "Bovine pepsin: the sequence of the first 65 amino acid residues
RL  (completing the sequence of the first 110 residues of bovine
RN  pepsinogen).";
RX  FEBS LETT. 60:133-136(1975).
CC  -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC  INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC  ALSO CLEAVED TO SOME EXTENT.
CC  -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC  EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR  PIR; A00983; PEBQ.
DR  PROSITE; PS00141; ASP-PROTEASE; 1.
DR  PFAM; PF00026; asp; 1.
DR  HSSP; P00790; IPSO.
KW  HYDROLASE; ASPARTYL PROTEASE; ZYMOMEN; DIGESTION; STOMACH;
KW  GASTRIC JUICE; GLYCOPROTEIN.
FT  PROPEP 1 41 ACTIVATION PEPTIDE.
FT  DOMAIN 1 17 PEPTIDE COMPRISING RESIDUES 1-17 INHIBITS
FT  THE ACTIVE ENZYME.
FT  PEP SIN A.
FT  CHAIN 42 >110
FT  ACT_SITE 77 77
FT  DISULFID 90 95
FT  NON_TER 110 110
SQ  SEQUENCE 110 AA; 12484 MW; 92BFB72 CRC32;

Query Match 44.9%; Score 61; DB 1; Length 110;
Best Local Similarity 57.1%; Pred. No. 3.57e+00;
Matches 8; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Db 73 TVIF-DTGSSNLNV 85
OY 44 TVLFHEPGSSSVW 57

RESULT 10
ID CDSA_ECOLI STANDARD; PRT; 249 AA.
AC P06466;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE PHOSPHATIDATE CYTIDYLYLTRANSFERASE (EC 2.7.7.41) (CDP-DIGLYCERIDE
DE SYNTHETASE) (CDP-DIGLYCERIDE PYROPHOSPHORYLASE) (CDP-DIACYLGLYCEROL
DE SYNTHASE) (CDS) (CTP:PHOSPHATIDATE CYTIDYLYLTRANSFERASE) (CDP-DAG
DE SYNTHASE).
GN CDSA OR CDS.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 86008268.
RA ICHO T., SPARROW C.P., RAETZ C.R.H.:
RT "Molecular cloning and sequencing of the gene for CDP-diglyceride
RL synthetase of Escherichia coli."
RN J. BIOL. CHEM. 260:12078-12083(1985).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MGI655;
RX BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,

```

```

RA MAU B., SHAO Y.:
RT "The complete genome sequence of Escherichia coli K-12."
RN SCIENCE 277:1453-1474(1997).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RA TAKEMOTO Y., MORI H., MURAYAMA N., KATAOKA K., YANO M., ITOH T.,
RA YAMAMOTO K., INOKUCHI H., MIKI T., HATADA E., FUKUDA R.,
RA ICHIHARA S., MIZUNO T., MAKINO K., NARATA A., YURA T., SAMPEI G.,
RA MIZOBUCHI K.:
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[4]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RA SCHRAMM S., DUNCAN M., ALLEN E., ARAUJO R., APARICIO A., CHUNG E.,
RA DAVIS K., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
RA LASHKARI D., LEW H., LIN D., NAMATH A., OEFNER P., ROBERTS D.,
RA DAVIS R.W.:
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[5]
RN CHARACTERIZATION.
RX MEDLINE; 86008269.
RA SPARROW C.P., RAETZ C.R.H.:
RT "Purification and properties of the membrane-bound CDP-diglyceride
RL synthetase from Escherichia coli."
RN J. BIOL. CHEM. 260:12084-12091(1985).
CC -!- CATALYTIC ACTIVITY: CTP + PHOSPHATIDATE -> PYROPHOSPHATE +
CC CDP-DIACYLGLYCEROL.
CC -!- PATHWAY: PHOSPHOLIPID BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CDS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11330; G145476; -.
DR EMBL; AE000127; G1786372; -.
DR EMBL; D83536; D1012617; -.
DR EMBL; U70214; G1552752; -.
DR PIR; A23898; SYECDG.
DR ECOGENE; EG10139; CDSA.
DR PFAM; PF01148; Cytidylyltrans; 1.
DR PFAM; PF01315; CDS; 1.
KW TRANSFERASE; NUCLEOTIDYLTRANSFERASE; PHOSPHOLIPID BIOSYNTHESIS;
KW TRANSMEMBRANE; INNER MEMBRANE.
FT TRANSMEM 20 40 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 154 174 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 228 248 POTENTIAL.
SQ SEQUENCE 249 AA; 27571 MW; B68B68C1 CRC32;

Query Match 44.9%; Score 61; DB 1; Length 249;
Best Local Similarity 58.3%; Pred. No. 3.57e+00;
Matches 7; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 67 VLFY-PGSAIIV 77
OY 45 VLFHEPGSSSVW 56

RESULT 11
ID PEPA_PIG STANDARD; PRT; 386 AA.
AC P00791;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

```

DE PEPsinogen A PRECURSOR (EC 3.4.23.1).
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88313677.
RA TSUKAGOSHI N., ANDO Y., TOMITA Y., UCHIDA R., TAKEMURA T.,
RA SAKAKI T., YAMAGATA H., UDAKA S., ICHIHARA Y., TAKAHASHI K.;
RT "Nucleotide sequence and expression in Escherichia coli of cDNA of
RT swine pepsinogen: involvement of the amino-terminal portion of the
RT activation peptide segment in restoration of the functional
RT protein.";
RL GENE 65:285-292(1988).
RN [2]
RP SEQUENCE OF 60-386.
RX MEDLINE; 74299591.
RA MORAVEK L., KOSTKA V.;
RT "Complete amino acid sequence of hog pepsin.";
RL FEBS LETT. 43:207-211(1974).
RN [3]
RP SEQUENCE OF 16-134.
RX MEDLINE; 74031413.
RA STEPANOV V.M., BARATOVA L.A., PUGACHEVA I.B., BELYANOVA L.P.,
RA REVINA L.P., TIMOKHINA E.A.;
RT "N-terminal sequence of swine pepsinogen and pepsin. The site of
RT pepsinogen activation.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 54:1164-1170(1973).
RN [4]
RP SEQUENCE OF 16-56.
RX MEDLINE; 69054241.
RA ONG E.B., PERLMANN G.E.;
RT "The amino-terminal sequence of porcine pepsinogen.";
RL J. BIOL. CHEM. 243:6104-6109(1968).
RN [5]
RP SEQUENCE OF 58-348.
RX MEDLINE; 75211282.
RA SEPULVEDA P., MARCINISZYN J.P. JR., LIU D., TANG J.;
RT "Primary structure of porcine pepsin. III. Amino acid sequence of a
RT cyanogen bromide fragment, CB2A, and the complete structure of
RT porcine pepsin.";
RL J. BIOL. CHEM. 250:5082-5088(1975).
RN [6]
RP PARTIAL SEQUENCE OF 1-26.
RX MEDLINE; 86059312.
RA ICHIHARA Y., SOGAWA K., TAKAHASHI K.;
RT "Isolation of human, swine, and rat prepepsinogens and calf
RT preprothymosin, and determination of the primary structures of their
RT NH2-terminal signal sequences.";
RL J. BIOCHEM. 98:483-492(1985).
RN [7]
RP ACTIVE SITE.
RX MEDLINE; 69283592.
RA BAYLISS R.S., KNOWLES J.R., WYBRANDT G.B.;
RT "An aspartic acid residue at the active site of pepsin. The isolation
RT and sequence of the heptapeptide.";
RL BIOCHEM. J. 113:377-386(1969).
RN [8]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE; 78077917.
RA ANDREEVA N.S., GUSTCHINA A.E., FEDOROV A.A., SHUTZKEVER N.E.,
RA VOLNOVA T.V.;
RT "X-ray crystallographic studies of pepsin.";
RL ADV. EXP. MED. BIOL. 95:23-31(1977).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE; 90317821.
RA COOPER J.B., KHAN G., TAYLOR G., TICKLE I.J., BLUNDELL T.L.;
RT "X-ray analyses of aspartic proteinases. II. Three-dimensional
RT structure of the hexagonal crystal form of porcine pepsin at 2.3-A
RT resolution.";
RL J. MOL. BIOL. 214:199-222(1990).
RN [10]

RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE; 91017500.
RA ABAD-ZAPATERO C., RYDEL T.J., ERICKSON J.;
RT "Revised 2.3 A structure of porcine pepsin: evidence for a flexible
RT subdomain.";
RL PROTEINS 8:62-81(1990).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE; 91278095.
RA SIELECKI A.R., FUJINAGA M., READ R.J., JAMES M.N.G.;
RT "Refined structure of porcine pepsinogen at 1.8-A resolution.";
RL J. MOL. BIOL. 219:671-692(1991).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RX MEDLINE; 92279205.
RA HARTSUCK J.E., KOELSCH G., REMINGTON S.J.;
RT "The high-resolution crystal structure of porcine pepsinogen.";
RL PROTEINS 13:1-25(1992).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-
CC -1- PPM: MINOR AMOUNTS OF THE ACTIVE ENZYME OCCUR WITH ALA-58 AT THE
CC AMINO END.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20920; G164602; -.
DR PIR; A00982; PEPG.
DR PIR; B22434; B22434.
DR PIR; J0307; J0307.
DR PDB; 3PEP; 15-APR-90.
DR PDB; 4PEP; 15-JAN-91.
DR PDB; 5PEP; 15-JUL-90.
DR PDB; 1PSA; 31-JAN-94.
DR PDB; 2PSG; 15-OCT-92.
DR PDB; 3PSG; 15-JAN-93.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR PFAM; PF00036; asp; 1.
KW HYDROLASE; ASPARTYL PROTEASE; DIGESTION; STOMACH; GASTRIC JUICE;
KW ZYMOGEN; SIGNAL; PHOSPHORYLATION; 3D-STRUCTURE.
FT SIGNAL 1 15
FT PROPEP 16 59 ACTIVATION PEPTIDE.
FT CHAIN 60 386 PEPSIN A.
FT MOD_RES 127 127 PHOSPHORYLATION.
FT ACT_SITE 91 91
FT ACT_SITE 274 274
FT DISULFID 104 109
FT DISULFID 265 269
FT DISULFID 309 342
FT CONFLICT 34 34 N -> D (IN REF. 3 AND 4).
FT CONFLICT 119 120 DS -> SD (IN REF. 2).
FT CONFLICT 128 128 Q -> E (IN REF. 2).
FT CONFLICT 289 289 MISSING (IN REF. 2 AND 5).
FT CONFLICT 314 314 S -> Q (IN REF. 5).
FT CONFLICT 323 323 N -> D (IN REF. 5).
FT STRAND 61 64
FT STRAND 66 68
FT TURN 69 71
FT TURN 72 79
FT STRAND 80 83
FT TURN 84 91
FT STRAND 92 93
FT TURN 97 99
FT STRAND 101 101

FT TURN 102 103
FT HELIX 107 110
FT TURN 111 111
FT STRAND 115 115
FT HELIX 117 119
FT STRAND 124 133
FT STRAND 138 150
FT TURN 151 152
FT STRAND 153 165
FT HELIX 169 172
FT TURN 173 173
FT STRAND 178 181
FT HELIX 185 187
FT HELIX 189 191
FT HELIX 195 201
FT TURN 202 203
FT STRAND 209 213
FT TURN 217 218
FT STRAND 219 219
FT STRAND 223 226
FT TURN 227 227
FT HELIX 231 233
FT STRAND 234 234
FT STRAND 239 242
FT STRAND 246 246
FT TURN 247 248
FT STRAND 249 258
FT TURN 259 260
FT STRAND 261 264
FT TURN 266 267
FT STRAND 269 273
FT TURN 275 276
FT STRAND 280 282
FT HELIX 284 288
FT TURN 290 295
FT TURN 296 296
FT STRAND 298 299
FT STRAND 305 307
FT HELIX 309 313
FT TURN 314 314
FT STRAND 318 322
FT TURN 323 324
FT STRAND 325 329
FT HELIX 331 334
FT STRAND 335 338
FT TURN 339 340
FT STRAND 341 344
FT STRAND 346 348
FT STRAND 352 353
...

Note: remainder of annotations omitted.

Query Match 44.9%; Score 61; DB 1; Length 386;
Best Local Similarity 57.1%; Pred. No. 3.57e+00;
Matches 8; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Db 87 TVIF-DTGSSNLWV 99
QY 44 TVLFEPGSSVWV 57
||| : ||| : ||

RESULT 12
ID PEP2_RABIT STANDARD; PRT; 387 AA.
AC P28713;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PEP2INOGEN II-4 PRECURSOR (EC 3.4.23.1) (PEPSINOGEN A).
OS ORYCTOLAGUS CONICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE: 91009127.
RA KAGEYAMA T., TANABE K., KOIWA O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. BIOL. CHEM. 265:17031-17038(1990).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: DIFFERENTIAL CLEAVAGE PHE-, LEU-.
CC -!- DEVELOPMENTAL STAGE: PEP2INOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -!- THE EXPRESSION OF PEP2INOGEN GENES IS REGULATED BY HORMONES AND
CC RELATED SUBSTANCES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR; D38302; D38302.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR PIR; P00026; asp; 1.
DR HSP; P00791; 1PSA.
KW HYDROLASE; ASPARTYL PROTEASE; DIGESTION; STOMACH; GASTRIC JUICE;
KW ZYMOGEN; SIGNAL; PHOSPHORYLATION; MULTIGENE FAMILY.
FT SIGNAL 1 15
FT PROPEP 16 59 ACTIVATION PEPTIDE.
FT CHAIN 60 387 PEP2IN II-4.
FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT DISULFID 106 111 BY SIMILARITY.
FT DISULFID 267 271 BY SIMILARITY.
FT DISULFID 310 343 BY SIMILARITY.
SQ SEQUENCE 387 AA; 42052 MW; 7CB48709 CRC32;
Query Match 44.9%; Score 61; DB 1; Length 387;
Best Local Similarity 57.1%; Pred. No. 3.57e+00;
Matches 8; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
Db 89 TVIF-DTGSSNLWV 101
QY 44 TVLFEPGSSVWV 57
||| : ||| : ||
RESULT 13
ID PEP2_RABIT STANDARD; PRT; 387 AA.
AC P27821;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PEP2INOGEN II-2/3 PRECURSOR (EC 3.4.23.1) (PEPSINOGEN A).
OS ORYCTOLAGUS CONICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91009127.
RA KAGEYAMA T., TANABE K., KOIWA O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. BIOL. CHEM. 265:17031-17038(1990).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: DIFFERENTIAL CLEAVAGE PHE-, LEU-.
CC -!- DEVELOPMENTAL STAGE: PEP2INOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -!- THE EXPRESSION OF PEP2INOGEN GENES IS REGULATED BY HORMONES AND
CC RELATED SUBSTANCES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; M59235; G165600; -
 PIR; C38302; C38302.
 PROSITE; PS00141; ASP_PROTEASE; 2.
 PFAM; PF00026; asp; 1.
 HSSP; P00790; 1PSO.
 HYDROLASE; ASPARTYL PROTEASE; DIGESTION; STOMACH; GASTRIC JUICE;
 ZMOGEN; SIGNAL; PHOSPHORYLATION; MULTIGENE FAMILY.
 SIGNAL 1 15
 FT PROPEP 16 59 ACTIVATION PEPTIDE.
 FT CHAIN 60 387 PEPsin II-2/3.
 FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
 FT ACT_SITE 93 93 BY SIMILARITY.
 FT ACT_SITE 276 276 BY SIMILARITY.
 FT DISULFID 106 111 BY SIMILARITY.
 FT DISULFID 267 271 BY SIMILARITY.
 FT DISULFID 310 343 BY SIMILARITY.
 SQ SEQUENCE 387 AA; 42100 MW; 048DC49C CRC32;

Query Match 44.9%; Score 61; DB 1; Length 387;
 Best Local Similarity 57.1%; Pred. No. 3.57e+00;
 Matches 8; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Db 89 TVIF-DTGSSNLWV 101
 ||| : ||| : |||
 Qy 44 TVLFHEPGSSVWV 57

RESULT 14
 ID PEP1_RABIT STANDARD; PRT; 387 AA.
 AC P28712;
 DT 01-DEC-1992 (REL. 24, CREATED)
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE PEPsinogen II-1 PRECURSOR (EC 3.4.23.1) (PEPSINOGEN A).
 OS ORYCTOLAGUS CUNICULUS (RABBIT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 91009127.
 RA KAGEYAMA T., TANABE K., KOIWA O.
 RT "Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucleotide sequences of cDNAs, molecular evolution, and gene expression during development.";
 RL J. BIOL. CHEM. 265:17031-17038(1990).
 CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE ALSO CLEAVED TO SOME EXTENT.
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
 CC -1- DEVELOPMENTAL STAGE: PEPsinogens IN GROUP I, II, AND III WHERE THE PREDOMINANT ZMOGENS AT LATE POSTNATAL STAGE.
 CC -1- THE EXPRESSION OF PEPsinogen GENES IS REGULATED BY HORMONES AND RELATED SUBSTANCES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 PIR; B38302; B38302.
 PROSITE; PS00141; ASP_PROTEASE; 2.
 PFAM; PF00026; asp; 1.
 HSSP; P00791; 1PSA.
 HYDROLASE; ASPARTYL PROTEASE; DIGESTION; STOMACH; GASTRIC JUICE;
 ZMOGEN; SIGNAL; PHOSPHORYLATION; MULTIGENE FAMILY.
 SIGNAL 1 15
 FT PROPEP 16 59 ACTIVATION PEPTIDE.
 FT CHAIN 60 387 PEPsin II-1.
 FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
 FT ACT_SITE 93 93 BY SIMILARITY.
 FT ACT_SITE 276 276 BY SIMILARITY.

FT DISULFID 106 111 BY SIMILARITY.
 FT DISULFID 267 271 BY SIMILARITY.
 FT DISULFID 310 343 BY SIMILARITY.
 SQ SEQUENCE 387 AA; 42070 MW; B8BE020B CRC32;

Query Match 44.9%; Score 61; DB 1; Length 387;
 Best Local Similarity 57.1%; Pred. No. 3.57e+00;
 Matches 8; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Db 89 TVIF-DTGSSNLWV 101
 ||| : ||| : |||
 Qy 44 TVLFHEPGSSVWV 57

RESULT 15
 ID PEP3_RABIT STANDARD; PRT; 387 AA.
 AC P27822;
 DT 01-AUG-1992 (REL. 23, CREATED)
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE PEPsinogen III PRECURSOR (EC 3.4.23.1) (PEPSINOGEN A).
 OS ORYCTOLAGUS CUNICULUS (RABBIT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 91009127.
 RA KAGEYAMA T., TANABE K., KOIWA O.
 RT "Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucleotide sequences of cDNAs, molecular evolution, and gene expression during development.";
 RL J. BIOL. CHEM. 265:17031-17038(1990).
 CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE ALSO CLEAVED TO SOME EXTENT.
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
 CC -1- DEVELOPMENTAL STAGE: PEPsinogens IN GROUP I, II, AND III WHERE THE PREDOMINANT ZMOGENS AT LATE POSTNATAL STAGE.
 CC -1- THE EXPRESSION OF PEPsinogen GENES IS REGULATED BY HORMONES AND RELATED SUBSTANCES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 PIR; E38302; E38302.
 PROSITE; PS00141; ASP_PROTEASE; 2.
 PFAM; PF00026; asp; 1.
 HSSP; P00791; 1PSA.
 HYDROLASE; ASPARTYL PROTEASE; DIGESTION; STOMACH; GASTRIC JUICE;
 ZMOGEN; SIGNAL; PHOSPHORYLATION; MULTIGENE FAMILY.
 SIGNAL 1 15
 FT PROPEP 16 59 ACTIVATION PEPTIDE.
 FT CHAIN 60 387 PEPsin III.
 FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
 FT ACT_SITE 93 93 BY SIMILARITY.
 FT ACT_SITE 276 276 BY SIMILARITY.
 FT DISULFID 106 111 BY SIMILARITY.
 FT DISULFID 267 271 BY SIMILARITY.
 FT DISULFID 310 343 BY SIMILARITY.
 SQ SEQUENCE 387 AA; 41969 MW; 898D3CAE CRC32;

Query Match 44.9%; Score 61; DB 1; Length 387;
 Best Local Similarity 57.1%; Pred. No. 3.57e+00;
 Matches 8; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Db 89 TVIF-DTGSSNLWV 101
 ||| : ||| : |||
 Qy 44 TVLFHEPGSSVWV 57

Db 89 TVIF-DTGSNLWV 101
||:| : ||:|
Qy 44 TVLFHEPGSSSVWV 57

Search completed: Thu Jul 8 18:10:27 1999
Job time : 7 secs.

THIS PAGE BLANK (USPTO)

MAQSEFH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsarch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:10:46 1999; MasPar time 8.08 Seconds

Tabular output not generated. 121.557 Million cell updates/sec

Title: >US-09-041-236-2
Description: (42-59) from US09041236.pep (4 of 45)
Perfect Score: 136

Sequence: 1 PHTVLFHEPGSSVWVG 18

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertebrate 14:sp.virus

Statistics: Mean 27.984; Variance 41.624; scale 0.672

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	136	100.0	666	4	075326 SEMAPHORIN L	1.00e-15
2	128	94.1	393	11	088371 SEMAPHORIN L (FRAGMENT	7.48e-14
3	88	64.7	633	14	064906 SIMILAR TO GENBANK ACC	5.09e-05
4	70	51.5	336	10	065453 ASPARTIC PROTEINASE LI	1.73e-01
5	69	50.7	428	5	018020 C15C8.3 PROTEIN.	2.65e-01
6	67	49.3	399	13	057477 CATHEPSIN D (EC 3.4.23	6.12e-01
7	66	48.5	439	3	004418 D9719.30P.	9.25e-01
8	65	47.8	182	2	033358 HYPOTHETICAL 19.4 KD P	1.39e+00
9	65	47.8	380	6	046498 PREGNANCY-ASSOCIATED G	1.39e+00
10	65	47.8	396	13	093428 CATHEPSIN D PRECURSOR.	1.39e+00
11	65	47.8	4836	11	088473 RJS.	1.39e+00
12	62	45.6	130	1	059275 130AA LONG HYPOTHETICA	4.65e+00
13	62	45.6	384	13	091322 PEPSINOGEN PRECURSOR.	4.65e+00
14	62	45.6	470	5	P91086 CODED FOR BY C. ELEGAN	4.65e+00
15	61	44.9	385	6	Q29080 PEPSINOGEN A PRECURSOR	6.89e+00
16	61	44.9	395	8	P92352 NADH DEHYDROGENASE SUB	6.89e+00
17	61	44.9	397	11	035647 CATHEPSIN E.	6.89e+00
18	61	44.9	433	10	004593 SIMILARITY TO BRASSICA	6.89e+00
19	61	44.9	442	5	093106 ASPARTIC PROTEASE (FRA	6.89e+00
20	61	44.9	509	10	Q39476 CYPROGIN.	6.89e+00

21	61	44.9	509	10	096383 ASPARTIC PROTEINASE.	6.89e+00
22	61	44.9	517	10	081654 SENESCENCE-ASSOCIATED	6.89e+00
23	61	44.9	1357	3	006832 CHROMOSOME IV COSMID 9	6.89e+00
24	60	44.1	63	1	029588 4-OXALOCROTONATE TAUTO	1.02e+01
25	60	44.1	95	14	084517 GENOME, PARTIAL SEQUEN	1.02e+01
26	60	44.1	202	1	027168 CONSERVED PROTEIN.	1.02e+01
27	60	44.1	255	10	043406 PUTATIVE ASPARTIC PROT	1.02e+01
28	60	44.1	283	10	082192 F8F22.8 PROTEIN.	1.02e+01
29	60	44.1	341	6	002728 PREGNANCY-ASSOCIATED G	1.02e+01
30	60	44.1	388	5	094123 ASPARTIC PROTEASE.	1.02e+01
31	60	44.1	398	13	087370 CATHEPSIN D (EC 3.4.23	1.02e+01
32	60	44.1	399	13	093458 CATHEPSIN D.	1.02e+01
33	60	44.1	419	11	009043 KIDNEY-DERIVED ASPARTI	1.02e+01
34	60	44.1	421	9	064302 A-PROTEIN.	1.02e+01
35	60	44.1	451	4	075072 FUKUTIN.	1.02e+01
36	60	44.1	486	10	038934 ASPARTIC PROTEINASE (F	1.02e+01
37	60	44.1	506	10	039311 F12F1.24 PROTEIN.	1.02e+01
38	60	44.1	506	10	065390 F44F1.1 PROTEIN.	1.02e+01
39	60	44.1	617	5	002259 PGSR-FAMILY PROTEIN.	1.02e+01
40	60	44.1	694	2	053212 ENDOGLUCANASE A (EC 3.	1.02e+01
41	60	44.1	1711	2	P96311 ASPARTYL PROTEASE NM1	1.49e+01
42	59	43.4	130	6	Q29452 HYPOTHETICAL 26.8 KD P	1.49e+01
43	59	43.4	244	2	053922 R12H7.2 PROTEIN.	1.49e+01
44	59	43.4	444	5	Q21966 HYPOTHETICAL 71.2 KD P	1.49e+01
45	59	43.4	630	3	059781	1.49e+01

ALIGNMENTS

RESULT 1					
ID	075326	PRELIMINARY;	PRT;	666	AA.
AC	075326;				
DT	01-NOV-1998	(TREMREL. 08, CREATED)			
DT	01-NOV-1998	(TREMREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L.				
GN	SEMAP.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;				
OC	CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98389619.				
RA	LANG E C., LIEBHART T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;				
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA				
RT	viruses."				
RL	GENOMICS 51:340-350(1998).				
DR	EMBL; AF030698; G3523115; -				
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;				

Query Match 100.0%; Score 136; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 1.00e-15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	74	PHTVLFHEPGSSVWVG	91
Qy	42	PHTVLFHEPGSSVWVG	59

RESULT 2					
ID	088371	PRELIMINARY;	PRT;	393	AA.
AC	088371;				
DT	01-NOV-1998	(TREMREL. 08, CREATED)			
DT	01-NOV-1998	(TREMREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L (FRAGMENT).				
GN	SEMAP.				
OS	MUS MUSCULUS (MOUSE).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;				
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98389619.				

QY	45	VLFEPGSSVWV	57
RESULT	6		
ID	057477	PRELIMINARY;	PRT; 399 AA.
AC	057477;		
DT	01-JUN-1998 (TREMBREL. 06, CREATED)		
DT	01-JUN-1998 (TREMBREL. 06, LAST SEQUENCE UPDATE)		
DT	01-AUG-1998 (TREMBREL. 07, LAST ANNOTATION UPDATE)		
DE	CATHEPSIN D (EC 3.4.23.5).		
OS	SPARUS AURATA (GILTHEAD SEA BREAM).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;		
OC	TELEOSTEI; EU TELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;		
OC	PERCIDEI; SPARIDAE; SPARUS.		


```

RN  SEQUENCE FROM N.A.
RP  TISSUE-OVARY;
RC  CARNEVALI O., BROOKS S., SUMPTER J., POLZONETTI-MAGNI A.;
RL  SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL: AF036319; G2687645; -
KW  PROSITE; PS00141; ASP-PROTEASE; 2.
SQ  SEQUENCE 399 AA; 43002 MW; 43002 MW; DIF5B87A CRC32;

Query Match 49.3%; Score 67; DB 13; Length 399;
Best Local Similarity 64.3%; Pred. No. 6.12e-01;
Matches 9; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db  92 TVLF-DTGSSNLWV 104
Qy  44 TVLFHPGSSSVWV 57

RESULT 7
ID  Q04418 PRELIMINARY; PRT; 439 AA.
AC  Q04418;
DT  01-NOV-1996 (TREMBLREL. 01, CREATED)
DT  01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT  01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE  D9719.30P.
DE  D9719.30.
GN  D9719.30.
OS  SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC  EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC  SACCHAROMYCETACEAE; SACCHAROMYCES.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  DIETRICH F.S., MOLLIGAN J., ALLEN E., ARAUJO R., AVILES E., BERNO A.,
RA  CARPENTER J., CHEN E., CHERRY J.M., CHONG E., DUNCAN M.,
RA  HUNICKE-SMITH S., HYMAN R., KOMP C., LASHKARI D., LEW H., LIN D.,
RA  MOSEDALE D., NAKAHARA K., NAMATH A., OEFNER P., OH C., PETEL F.X.,
RA  ROBERTS D., SCHRAMM S., SCHROEDER M., SHOGEN T., SHROFF N.,
RA  WINANT A., YELTON M., BOTSTEIN D., DAVIS R.W.;
RL  SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN  [2]
RP  SEQUENCE FROM N.A.
RA  DIETRICH F.S.;
RL  SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN  [3]
RP  SEQUENCE FROM N.A.
RA  JIA Y., CHERRY J.M.;
RL  SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL: U33057; G927794; -
SQ  SEQUENCE 439 AA; 50106 MW; 362B6C8C CRC32;

Query Match 48.5%; Score 66; DB 3; Length 439;
Best Local Similarity 55.6%; Pred. No. 9.25e-01;
Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Db  129 NTPLFAIEGASGTWVG 146
Qy  43 HTVLFHE-PGSSSVWVG 59

RESULT 8
ID  Q33358 PRELIMINARY; PRT; 182 AA.
AC  Q33358;
DT  01-JAN-1998 (TREMBLREL. 05, CREATED)
DT  01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT  01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE  HYPOTHETICAL 19.4 KD PROTEIN.
GN  MTCY20G10.03.
OS  MYCOBACTERIUM TUBERCULOSIS.
OC  BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIACEAE;
OC  ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-H37RV;

BROWN D., CHURCHER C.M.;
RL  SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN-H37RV;
RA  PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL  SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN-H37RV;
RX  MEDLINE; 96181548.
RA  PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA  BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA  COLE S.T.;
RT  "An integrated map of the genome of the tubercle bacillus,
RT  Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT  leprae.";
RL  PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
DR  EMBL: Z97831; E329664; -
KW  HYPOTHETICAL PROTEIN.
SQ  SEQUENCE 182 AA; 19401 MW; 2E881D7F CRC32;

Query Match 47.8%; Score 65; DB 2; Length 182;
Best Local Similarity 61.5%; Pred. No. 1.39e+00;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db  11 LLFYEPGASWYV 23
Qy  45 VLFHPGSSSVWV 57

RESULT 9
ID  Q46498 PRELIMINARY; PRT; 380 AA.
AC  Q46498;
DT  01-JUN-1998 (TREMBLREL. 06, CREATED)
DT  01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT  01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE  PREGNANCY-ASSOCIATED GLYCOPROTEIN 10.
GN  PAG10.
OS  BOS TAURUS (BOVINE).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE-PLACENTA;
RA  XIE S., GREEN J., BIXBY J., SZAFRANSKA B., DEMARTINI J.C., HECHT S.,
RA  ROBERTS R.M.;
RL  PROC. NATL. ACAD. SCI. U.S.A. 0-0-0(1998).
DR  EMBL: AF020512; G2921124; -
SQ  SEQUENCE 380 AA; 42737 MW; CCB09CCB CRC32;

Query Match 47.8%; Score 65; DB 6; Length 380;
Best Local Similarity 69.2%; Pred. No. 1.39e+00;
Matches 9; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db  87 VLF-DTGSSSLWV 98
Qy  45 VLFHPGSSSVWV 57

RESULT 10
ID  Q93428 PRELIMINARY; PRT; 396 AA.
AC  Q93428;
DT  01-NOV-1998 (TREMBLREL. 08, CREATED)
DT  01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT  01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE  CATHEPSIN D PRECURSOR.
OS  CHIONODRACO HAMATUS.
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC  TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;
OC  NOTOTHENIOIDEI; CHANNICHTHYDAE; CHIONODRACO.
RN  [1]
RP  SEQUENCE FROM N.A.
```

```
RC TISSUE=LIVER;
RA CAPASSO C., LEES W.E., CAPASSO A., SCUDIERO R., CARGINALE V.,
RA KILLE P., KAY J., PARISI E.;
RT "Isolation, characterization and sequencing of a cathepsin D from the
RT liver of the antarctic icefish Chionodraco hamatus.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AJ007878; E1313621; -.
DR EMBL: AF000006; D1031629; -.
SQ SEQUENCE 396 AA; 42662 MW; 55D574FC CRC32;

Query Match 47.8%; Score 65; DB 13; Length 396;
Best Local Similarity 56.3%; Pred. No. 1.39e+00;
Matches 9; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

Db 88 PFTVVF-DTGSSNLWV 102
QY 42 PHTVLFHEPGSSVWV 57

RESULT 11
ID 088473 PRELIMINARY; PRT; 4836 AA.
AC 088473;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RJS.
GN RJS.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE; 98356175.
RA LEHMAN A.L., NAKATSU Y., CHING A., BRONSON R.T., OAKLEY R.J.,
RA KEIPER-HYRNKO N., FINGER J.N., DURHAM-PIERRE D., HORTON D.B.,
RA NEWTON J.M., LYON M.F., BRILLIANT M.H.;
RT "A very large novel protein with diverse functional motifs is
RT deficient in rjs (runt), jerky, sterile" mice."
RL PROC. NATL. ACAD. SCI. U.S.A. 95:9436-9441(1998).
DR EMBL: AF061529; G3414809; -.
SQ SEQUENCE 4836 AA; 527362 MW; 12B62FC2 CRC32;

Query Match 47.8%; Score 65; DB 11; Length 4836;
Best Local Similarity 50.0%; Pred. No. 1.39e+00;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 237 PEASLFDSTVSSVWL 252
QY 42 PHTVLFHEPGSSVWV 57

RESULT 12
ID 059275 PRELIMINARY; PRT; 130 AA.
AC 059275;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE 130AA LONG HYPOTHETICAL PROTEIN.
GN PH1574.
OS PYROCOCCUS HORTKOSHII.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; PYROCOCCACEAE; PYROCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOVAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
```

```
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete Sequence and Gene Organization of the Genome of a
RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 5:55-76(1998).
DR EMBL: AF000006; D1031629; -.
SQ SEQUENCE 130 AA; 13603 MW; 20EF15A1 CRC32;

Query Match 45.6%; Score 62; DB 1; Length 130;
Best Local Similarity 53.3%; Pred. No. 4.65e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 61 PTTIGFFNPGVTSVW 75
QY 42 PHTVLFHEPGSSVWV 56

RESULT 13
ID 091322 PRELIMINARY; PRT; 384 AA.
AC 091322;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PEPsinogen precursor.
GN PEPsinogen.
OS RANA CATESBEIANA (BULL FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC NEOBATRACHIA; RANOIDEA; RANIDAE; RANINAE; RANA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92042186.
RA YAKABE E., TANJI M., ICHINOSE M., GOTO S., MIKI K., KUROKAWA K.,
RA ITO H., KAGEYAMA T., TAKAHASHI K.;
RT "Purification, characterization, and amino acid sequences of
RT pepsinogens and pepsins from the esophageal mucosa of bullfrog (Rana
RT catesbeiana).";
RL J. BIOL. CHEM. 266:22436-22443(1991).
DR EMBL: M73750; G213688; -.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR PFAM; PF00026; asp_1.
KW SIGNAL; HYDROLASE; ASPARTYL PROTEASE.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 53 384 PEPsin.
SQ SEQUENCE 384 AA; 41764 MW; D2AA757F CRC32;

Query Match 45.6%; Score 62; DB 13; Length 384;
Best Local Similarity 61.5%; Pred. No. 4.65e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 82 VLF-DTGSSNLWV 93
QY 45 VLFHEPGSSVWV 57

RESULT 14
ID P91086 PRELIMINARY; PRT; 470 AA.
AC P91086; O02545;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CODED FOR BY C. ELEGANS CDNA YK134H12.5.
GN C24A11.4 OR F55A12.9.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
```

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA MAGGI L.;
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA PAULEY A., GATTUNG S.;
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U80443; G1703624; -;
DR EMBL; AF003130; G2088669; -;
SQ SEQUENCE 470 AA: 54308 MW; 86416042 CRC32;

Query Match 45.6%; Score 62; DB 5; Length 470;
Best Local Similarity 37.5%; Pred. No. 4.65e+00;

Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 452 PRTLFLYPPDAATWI 467

Qy 42 PTVLFHEPGSSVWV 57

RESULT 15
ID Q29080 PRELIMINARY; PRT; 385 AA.
AC Q29080;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE PERSINOGEN A. PRECURSOR.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89174702.
RA LIN X.L., WONG R.N.S., TANG J.;
RT "Synthesis, purification, and active site mutagenesis of recombinant
porcine pepsinogen.";
RL J. BIOL. CHEM. 264:4482-4489(1989).
DR EMBL; J04601; G164604; -;
DR PROSITE: PS00141; ASP_PROTEASE; 2.
DR PFAM: PF00026; asp; 1.
KW SIGNAL; HYDROLASE; ASPARTYL PROTEASE.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 385 POTENTIAL.
FT CHAIN 60 385 POTENTIAL.
SQ SEQUENCE 385 AA: 41310 MW; 684154DB CRC32;

Query Match 44.9%; Score 61; DB 6; Length 385;
Best Local Similarity 57.1%; Pred. No. 6.89e+00;

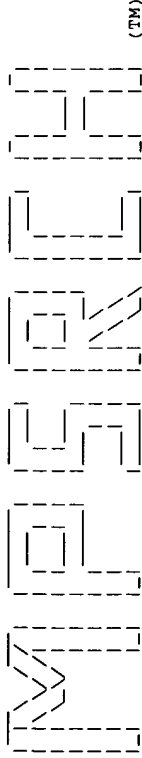
Matches 8; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Db 87 TVIF-DTGSSNLWV 99

Qy 44 TVLFHEPGSSVWV 57

Search completed: Thu Jul 8 18:10:58 1999
Job time : 12 secs.

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:14:12 1999; MasPar time 10.97 Seconds
Tabular output not generated. 19.392 Million cell updates/sec

Title: >US-09-041-236-2
Description: (62-71) from US09041236.pap (5 of 45)
Perfect Score: 83
Sequence: 1 KVVYLFDFPEG 10

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 18.168; Variance 61.134; scale 0.297

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	67	80.7	684	36	W49873 Thermotoga maritima M	5.66e+00
2	58	69.9	88	36	W71644 Vaccinia virus transl	4.52e+01
3	58	69.9	88	5	R25832 K3L	4.52e+01
4	58	69.9	88	37	W71365 K3L translation facto	4.52e+01
5	58	69.9	945	39	W73624 Human secreted protei	4.52e+01
6	52	62.7	439	36	W70472 South African Arbovir	1.72e+02
7	52	62.7	439	36	W70486 Sindbis virus EL prot	1.72e+02
8	52	62.7	439	36	W70479 Girdwood S.A.virus E1	1.72e+02
9	51	61.4	51	19	W03603 Rat serotonin 1c GPR	2.14e+02
10	51	61.4	82	1	R06543 Post transfusion non-	2.14e+02
11	51	61.4	280	12	R53760 sACS anti-terminator.	2.14e+02
12	51	61.4	411	36	W72745 Drosophila kinesin N-	2.14e+02
13	51	61.4	418	1	P90549 Serotonin 5HT1c recep	2.14e+02
14	51	61.4	441	36	W72744 Drosophila kinesin N-	2.14e+02
15	51	61.4	453	12	R60697 Fragment of the rat s	2.14e+02
16	51	61.4	460	37	W77105 Rat 5-HT2C serotonin	2.14e+02

17	51	61.4	460	1	P92111 Serotonin 5HT1c recep	2.14e+02
18	51	61.4	460	37	W77113 Rat 5-HT2C serotonin	2.14e+02
19	51	61.4	460	37	W77112 Rat 5-HT2C serotonin	2.14e+02
20	51	61.4	975	36	W72746 Drosophila kinesin.	2.14e+02
21	51	61.4	2273	20	R98811 Erysiphe graminis ace	2.14e+02
22	50	60.2	71	31	W28085 Staphylococcus aureus	2.66e+02
23	50	60.2	450	37	W73137 Bax inhibitor BI-2.	2.66e+02
24	50	60.2	532	16	R82937 Mouse Kv1.7 voltage-g	2.66e+02
25	50	60.2	878	31	W56116 Microbispora thermoro	2.66e+02
26	49	59.0	77	28	W26745 S. carnosus nitrate r	3.30e+02
27	49	59.0	263	30	W56139 Open reading frame 4	3.30e+02
28	49	59.0	455	28	W41076 P. furiosus ADPHK pro	3.30e+02
29	49	59.0	467	28	W41077 T. litoralis ADPHK pr	3.30e+02
30	49	59.0	566	37	W82559 A. thaliana homologou	3.30e+02
31	49	59.0	900	2	R06345 Viral enhancing facto	3.30e+02
32	49	59.0	3457	12	R62504 Large polyprotein seq	3.30e+02
33	48	57.8	320	29	W55501 H. pylori ORF 01gel16	4.08e+02
34	48	57.8	324	29	W55384 H. pylori ORF 11gel103	4.08e+02
35	48	57.8	578	25	W26607 Thermostable DNA poly	4.08e+02
36	48	57.8	608	25	W26606 Thermostable DNA poly	4.08e+02
37	48	57.8	872	25	W26605 Thermoanaerobacter th	4.08e+02
38	48	57.8	1187	28	W42990 Amino acid sequence o	4.08e+02
39	48	57.8	1188	28	W42991 Amino acid sequence o	4.08e+02
40	48	57.8	1216	20	W10685 Human tyrosine dephos	4.08e+02
41	47	56.6	185	39	W82004 Human adult testis se	5.05e+02
42	47	56.6	216	31	W46293 Streptococcus pneumon	5.05e+02
43	47	56.6	642	9	R49826 Cephalosporin C #1.	5.05e+02
44	47	56.6	764	38	W81784 P. putida cis/trans i	5.05e+02
45	47	56.6	1132	17	R77417 Human cell cycle prot	5.05e+02

ALIGNMENTS

RESULT 1
ID W49873 standard; Protein: 684 AA.

AC W49873;
DT 21-DEC-1998 (first entry)
DE Thermotoga maritima MSB8-6GB4 glycosidase.
KW Glycosidase; MSB8-6GB4; thermostable enzyme; oligosaccharide;
KW glucose; sugar; baking; textile; detergent; beta-galactosidase.
OS Thermotoga maritima strain MSB8-6GB4.
PN W09824799-A1.
PD 11-JUN-1998.
PF 08-DEC-1997; U22623.
PR 10-OCT-1997; US-949026.
PR 06-DEC-1996; US-056916.
PA (DIVE-) DIVERSA CORP.
PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
DR WPI: 98-362407/31.
DR N-PSDB; V36922.
PT Glycosidase enzymes from organisms of the genera Staphylothermus,
PT Pyrococcus and Thermococcus - for deriving sugar from
PT oligosaccharides, useful in the e.g. food processing, textile or
PT baking industries
PS Claim 1: Fig 16a-c; 92pp; English.
CC This is the amino acid sequence of glycosidase MSB8-6GB4, deduced
CC from a polynucleotide (see V36922) of clone 6GB4 of Thermotoga
CC maritima MSB8 that grows optimally at 85 degC in high salt medium.
CC The invention provides 18 polynucleotides (see V36907-24) coding
CC for thermostable glycosidases (see W49858-75) having glucosidase,
CC alpha-galactosidase, beta-galactosidase, beta-mannosidase,
CC beta-mannanase, endoglucanase or pullulanase activity. Vectors and
CC host cells are also claimed. A method is provided for producing
CC generating glucose from soluble cell oligosaccharides comprises
CC contacting a sample (selected from dairy products, fruit juice,
CC detergent, textile, guar gum, animal feed, plant biomass or waste
CC product) containing oligosaccharides (selected from maltose,
CC cellobiose, lactose, sucrose, raffinose, stachyose, polysaccharides
CC and pullulan) with one of the claimed glycosidases such that glucose
CC is produced.
SQ Sequence 684 AA;

```

Query Match      80.7%; Score 67; DB 36; Length 684;
Best Local Similarity 77.8%; Pred. No. 5.56e+00;
Matches          7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db      441 rlyldfpe 449
QY      62 KVLDFDPE 70

RESULT 2
ID W71644 standard; Protein; 88 AA.
AC W71644;
DT 18-JAN-1999 (first entry)
DE Vaccinia virus translation factor K3L.
KW Vector; gene therapy; vaccine; ALVAC; translation factor; K3L; E3L.
OS Vaccinia virus.
PN W09840500-A1.
PD 17-SEP-1998.
PF 25-FEB-1998; U03710.
PR 12-MAR-1997; US-815809.
PA (UYAR-) UNIV ARIZONA STATE.
PA (VIRO-) VIROGENETICS CORP.
PI Cox WI, Gettig RR, Goebel SJ, Jacobs BL, Paoletti E,
PI Pincus SE, Tartaglia J;
DR WPI: 98-520819/44.
DR N-PSDB; V58245.
DR Enhancing expression of nucleic acids in cells - by using modified
PT vectors which comprise the nucleic acid and also nucleic acid
PT encoding a translation factor
PT Example 1; Fig 6A-B; 90pp; English.
PS This polypeptide comprises the K3L translation factor of vaccinia
CC virus. K3L and E3L (see W30564) play a role in the regulation of
CC viral translation. Both are capable of inhibiting the action of
CC a cellular protein kinase (PKR) which, when activated by double
CC stranded (ds) RNA, phosphorylate the translational initiation
CC factor eIF-2alpha, leading to an inhibition of initiation of mRNA
CC translation. Novel vectors of the invention, including ALVAC
CC vectors (see V58245) include K3L and E3L sequences, and are used
CC for enhancing expression of gene products that they encode. The
CC effect of the translation factors is to increase the effective
CC concentration of dsRNA. An up-regulation of foreign gene
CC expression can have a profound effect on the induction of a
CC therapeutic or immunological response in a host administered or
CC inoculated with recombinants derived from these new vectors,
CC thereby providing an enhanced immunological, e.g. protective,
CC response, or an enhanced therapeutic response.
SQ Sequence 88 AA;

Query Match      69.9%; Score 58; DB 36; Length 88;
Best Local Similarity 55.6%; Pred. No. 4.52e+01;
Matches          5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db      28 lylfdyphs 36
QY      63 VYLFDFPEG 71

RESULT 3
ID R25832 standard; protein; 88 AA.
AC R25832;
DT 20-JAN-1993 (first entry)
DE Vaccinia virus; VV; Copenhagen strain; VP872; pK3Lqpt; ORF; PCR.
OS Vaccinia virus.
PN W09212240-A.
PD 23-JUL-1992.
PF 06-JAN-1992; U00087.
PR 07-JAN-1991; US-638080.
PR 16-DEC-1991; US-805567.
PA (VIRO-) VIROGENETICS CORP.
PI Paoletti E, Tartaglia J;
DR WPI: 92-286664/32.

```

```

PT New recombinant pox-virus with enhanced sensitivity to interferon
PT - lacks open reading frame conferring interferon resistance and
PT vaccines contg. the virus may be inactivated by administering
PT interferon
PS Disclosure; Fig 1; 32pp; English.
CC This sequence is encoded by the K3L gene of vaccinia virus (VV).
CC Within the scope of the invention the expression of this protein is
CC disrupted. The strain of VV used was the Copenhagen strain. The
CC mutant VP872 was derived from this by specific deletion of the K3L
CC open reading frame (ORF). The K3L ORF encodes a 10.5 kD protein.
CC VP872 was generated using the deletion plasmid pK3Lqpt. The 5' and 3'
CC sequences relative to the K3L ORF were derived by PCR. (See also
CC Q26737-40). The deletion of K3L from the VV resulted in an enhanced
CC sensitivity of viral-induced protein synthesis to interferon.
SQ Sequence 88 AA;

Query Match      69.9%; Score 58; DB 5; Length 88;
Best Local Similarity 55.6%; Pred. No. 4.52e+01;
Matches          5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db      28 lylfdyphs 36
QY      63 VYLFDFPEG 71

RESULT 4
ID W71365 standard; Protein; 88 AA.
AC W71365;
DT 15-FEB-1999 (first entry)
DE K3L translation factor.
DE Vector; gene therapy; vaccine; ALVAC; K3L; translation factor.
KW Vaccinia virus.
PN W09840501-A1.
PD 17-SEP-1998.
PF 13-FEB-1998; U02669.
PR 12-MAR-1997; US-816155.
PA (VIRO-) VIROGENETICS CORP.
PI Cox WI, Gettig RR, Martinez H, Paoletti E, Pincus SE,
PI Tartaglia J;
DR WPI: 98-520820/44.
DR N-PSDB; V60353.
DR Enhancing expression of nucleic acids in cells - by using modified
PT vectors which comprise nucleic acid and also nucleic acid encoding
PT transcription factor and optionally translation factor
PT Example 2; Fig 8; 102pp; English.
PS This polypeptide comprises the K3L translation factor of vaccinia
CC virus. K3L and E3L (see W30571) play a role in the regulation of
CC viral translation. Both are capable of inhibiting the action of
CC a cellular protein kinase (PKR) which, when activated by double
CC stranded (ds) RNA, phosphorylate the translational initiation
CC factor eIF-2alpha, leading to an inhibition of initiation of mRNA
CC translation. Novel vectors are provided for enhanced expression of
CC at least 1 first nucleic acid molecule (NAM) in a cell of
CC particular phenotype. The vector (e.g. NIVAC or ALVAC) is modified
CC to comprise the first NAM and at least 1 second NAM encoding a
CC transcription factor (TF), or a TF and a translation factor such as
CC K3L and E3L, where there is co-temporal expression of the first
CC and second NAMS with respect to the phenotype of the cell, and
CC where expression of the second NAM enhances expression of the first
CC NAM by enhancing transcription or translation and translation.
CC Also claimed is a method for increasing expression of at least 1
CC first NAM by such a vector. The vectors can be used for increasing
CC expression of e.g. an epitope of interest, a biological response
CC modulator, a growth factor, a recognition sequence, a therapeutic
CC gene or a fusion protein.
SQ Sequence 88 AA;

Query Match      69.9%; Score 58; DB 37; Length 88;
Best Local Similarity 55.6%; Pred. No. 4.52e+01;
Matches          5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db      28 lylfdyphs 36
QY      63 VYLFDFPEG 71

```

QY 63 VYLFDFPEG 71

RESULT 5
ID W73624; standard; protein; 945 AA.
AC W73624;
DE 23-MAR-1999 (first entry)
DT Human secreted protein clone aw92_1.
KW Secreted protein; human; nutritional supplements; immune stimulant;
KW immune suppressor; hematopoiesis regulator; tissue growth; chemotaxis;
KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
KW cadherin/tumour invasion suppressor.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 493 /note= "unspecified amino acid"
FT Misc_difference 494 /note= "unspecified amino acid"
FT Misc_difference 921 /note= "unspecified amino acid"
FT W0985614-A2.
PN 10-DEC-1998.
PD 01-JUN-1998; U11210.
PE 29-MAY-1998; US-087255.
PR 04-JUN-1997; US-868696.
PR 04-JUN-1997; US-868697.
PR 04-JUN-1997; US-868698.
PR 04-JUN-1997; US-868699.
PR 04-JUN-1997; US-868899.
PR 04-JUN-1997; US-868900.
PR 04-JUN-1997; US-869191.
PR 04-JUN-1997; US-869192.
PR 04-JUN-1997; US-869193.
PR 04-JUN-1997; US-869194.
PR (GEM) GENETICS INST INC.
PI Agostino MJ, Fecthel K, Howes SH, Jacobs K, Lavallie ER,
PI McCoy JM, Racie LA, Spaulding V, Treacy M,
DR WPI: 99-059912/05.
DR N-PSDB; V53742.
PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
PT retina, adult placenta or adult uterus cDNA libraries
PS Claim 13: Page 76-79: 127pp; English.
CC This sequence represents a human secreted protein of the invention.
CC The DNA encoding this sequence was isolated from a human adult ovary cDNA
CC library, and was designated clone aw92_1. The DNAs and proteins
CC are predicted to have biological activities which would make them
CC suitable for treating, preventing or ameliorating medical conditions in
CC humans and animals, although no supporting data is given. Suggested
CC activities include nutritional sources or supplements, immune
CC stimulating or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The DNAs
CC are also stated to be useful for gene therapy. A host cell transfected
CC with the DNA, or its subfragments and variants is useful for recombinant
CC production of the human secreted protein clones.
SQ Sequence 945 AA;

Query Match 69.98; Score 58; DB 39; Length 945;
Best Local Similarity 60.08; Pred. No. 4.52e+01;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 707 klfydlppeg 716
QY 62 KVLDFDFPEG 71

RESULT 6
ID W70472 standard; Protein; 439 AA.
AC W70472;

DT 10-DEC-1998 (first entry)
DE South African Arbovirus strain No. 86_E1 protein sequence.
KW Bone marrow cell; alphavirus; South African Arbovirus strain No. 86;
KW S.A.AR86; reverse transcriptase-PCR; growth hormone; growth factor;
KW interleukin; cytokine; chemokine; enzyme; ribozyme; El;
KW antisense oligonucleotide.
OS South african arbovirus.
PN W09836779-A2.
PD 27-AUG-1998.
PE 18-FEB-1998; U02945.
PR 19-FEB-1997; US-801263.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Davis NL, Johnston RE, Simpson DA;
DR WPI: 98-495361/42.
DR N-PSDB; V33473.
PT Expression of heterologous RNA in bone marrow cells - using a
PT recombinant alphavirus comprising a promoter operable in bone marrow
PT cells operably associated with heterologous RNA.
PS Example 2: Fig 2B; 68pp; English.
CC The present sequence represents the South African Arbovirus strain
CC No. 86 (S.A.AR86) E1 protein sequence. The E1 protein is
CC encoded by the S.A.AR86 cDNA (V33473) as part of a structural
CC polypeptide. The invention provides a method for introducing and
CC expressing heterologous RNA in bone marrow cells using alphavirus
CC vectors. The S.A.AR86 is an alphavirus whose cDNA sequence was
CC determined from uncloned reverse transcriptase-PCR reaction fragments
CC amplified from the virion RNA. The inventors claim the transformed
CC bone marrow cells are useful for expressing a protein or peptide
CC suitable for protecting the subject against a disease such as a
CC microbial, bacterial, protozoal, parasitic or viral disease. The
CC transformed bone marrow cells are also claimed to be useful for
CC expressing proteins and peptides such as hormones, growth hormones,
CC growth factors, interleukins, cytokines, chemokines, enzymes, ribozymes
CC or antisense oligonucleotides.
SQ Sequence 439 AA;
Query Match 62.78; Score 52; DB 36; Length 439;
Best Local Similarity 75.08; Pred. No. 1.72e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 184 vynydfpe 191
QY 63 VYLFDFPE 70
RESULT 7
ID W70486 standard; Protein; 439 AA.
AC W70486;
DT 10-DEC-1998 (first entry)
DE Sindbis virus E1 protein sequence.
KW Bone marrow cell; alphavirus; Sindbis virus; growth hormone;
KW reverse transcriptase-PCR; growth factor; interleukin; cytokine;
KW chemokine; enzyme; ribozyme; antisense oligonucleotide; El.
OS Sindbis virus.
PN W09836779-A2.
PD 27-AUG-1998.
PE 18-FEB-1998; U02945.
PR 19-FEB-1997; US-801263.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Davis NL, Johnston RE, Simpson DA;
DR WPI: 98-495361/42.
DR N-PSDB; V33476.
PT Expression of heterologous RNA in bone marrow cells - using a
PT recombinant alphavirus comprising a promoter operable in bone marrow
PT cells operably associated with heterologous RNA.
PS Example 6: Page -; 68pp; English.
CC The present sequence represents the Sindbis virus E1 protein
CC sequence. The E1 protein is encoded by the Sindbis virus
CC cDNA (V33476) as part of a structural polypeptide. The invention
CC provides a method for introducing and expressing heterologous RNA in
CC bone marrow cells using alphavirus vectors. The inventors claim the
CC transformed bone marrow cells are useful for expressing a protein or
CC peptide suitable for protecting the subject against a disease such as

CC a microbial, bacterial, protozoal, parasitic or viral disease. The
 CC transformed bone marrow cells are also claimed to be useful for
 CC expressing proteins and peptides such as hormones, growth hormones,
 CC growth factors, interleukins, cytokines, chemokines, enzymes, ribozymes
 CC or antisense oligonucleotides.
 CC Note: The present sequence was not provided by the inventors in the
 CC specification, but was decoded using the Sindbis virus cDNA sequence.
 SQ Sequence 439 AA;

Query Match 62.7%; Score 52; DB 36; Length 439;
 Best Local Similarity 75.0%; Pred. No. 1.72e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 184 vynydfpe 191
 ||:||||
 QY 63 VYLFDFPE 70

RESULT 8
 ID W70479 standard; Protein: 439 AA.
 AC W70479;
 DE Girdwood S.A.virus E1 protein sequence.
 KW Bone marrow cell; alphavirus; Girdwood S.A.virus; growth hormone;
 KW reverse transcriptase-PCR; growth factor; interleukin; cytokine;
 KW chemokine; enzyme; ribozyme; antisense oligonucleotide; E1.
 OS Girdwood S.A.virus.
 PN WO9836779-A2.
 PD 27-AUG-1998.
 PF 18-FEB-1998; U02945.
 PR 19-FEB-1997; US-801263.
 PA (UNNC-) UNIV NORTH CAROLINA.
 PI Davis NL, Johnston RE, Simpson DA;
 DR WPI: 98-495361/42.
 DR N-PSDB: V33474.
 PT Expression of heterologous RNA in bone marrow cells - using a
 PT recombinant alphavirus comprising a promoter operable in bone marrow
 PT cells operably associated with heterologous RNA.
 PS Example 2; Fig 4B; 68pp; English.
 CC The present sequence represents the Girdwood S.A.virus E1 protein
 CC sequence. The E1 protein is encoded by the Girdwood S.A.virus
 CC cDNA (V33474) as part of a structural polyprotein. The invention
 CC provides a method for introducing and expressing heterologous RNA in
 CC bone marrow cells using alphavirus vectors. The Girdwood S.A.virus is
 CC an alphavirus whose cDNA sequence was determined from uncloned reverse
 CC transcriptase-PCR reaction fragments amplified from the virion RNA.
 CC The inventors claim the transformed bone marrow cells are useful for
 CC expressing a protein or peptide suitable for protecting the subject
 CC against a disease such as a microbial, bacterial, protozoal, parasitic
 CC or viral disease. The transformed bone marrow cells are also claimed
 CC to be useful for expressing proteins and peptides such as hormones,
 CC growth hormones, growth factors, interleukins, cytokines, chemokines,
 CC enzymes, ribozymes or antisense oligonucleotides.
 SQ Sequence 439 AA;

Query Match 62.7%; Score 52; DB 36; Length 439;
 Best Local Similarity 75.0%; Pred. No. 1.72e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 184 vynydfpe 191
 ||:||||
 QY 63 VYLFDFPE 70

RESULT 9
 ID W03603 standard; peptide; 51 AA.
 AC W03603;
 DE 17-DEC-1996 (first entry)
 DE Rat serotonin 1c GPR N-terminal sequence.
 KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
 KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
 KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
 KW odorant; cytomagalovirus; serotonergic.

OS Rattus rattus.
 PN US508384-A.
 PD 16-APR-1996.
 PF 10-SEP-1992; 943236.
 PR 10-SEP-1992; US-943236.
 PR 09-SEP-1993; US-118270.
 PA (UINY) UNIV NEW YORK STATE.
 PI Murphy RB, Schuster DI;
 DR WPI: 96-208785/21.
 PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.
 PT for treating schizophrenia
 PS Disclosure; Fig 8B(1); 184pp; English.
 CC Proteins W04657-W02720 represent a range of G-protein coupled receptor
 CC (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
 CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
 CC odorant, cytomagaloviral and other GPR proteins. The peptides
 CC W03578-W03651 represent the N-terminal fragments of the above proteins.
 CC The receptor proteins were used to design polypeptides, pref. based on
 CC the transmembrane domains, for use in G-protein coupled receptor ligand
 CC binding assays. The polypeptide fragments retain biological activity
 CC such as binding a GPR ligand or modulating GPR ligand binding to a GPR
 CC (see W02747-W02999 for examples of polypeptide fragments).
 CC The polypeptide fragments can be used in compositions for treating
 CC subjects suffering from a pathology related to a GPR abnormality e.g. a
 CC psychotic disorder such as schizophrenia.
 SQ Sequence 51 AA;

Query Match 61.4%; Score 51; DB 19; Length 51;
 Best Local Similarity 71.4%; Pred. No. 2.14e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 44 lfqfpgd 50
 ||:||||
 QY 65 LFDPEEG 71

RESULT 10
 ID R06543 standard; protein; 82 AA.
 AC R06543;
 DT 04-JAN-1991 (first entry)
 DE Post transfusion non-A non-B hepatitis virus gene product of
 DE Clobe MYSB-1
 KW YS1; YS2; NANB hepatitis.
 OS Non-A non-B hepatitis virus.
 PN J02186990-A.
 PD 23-JUL-1990.
 PF 10-JAN-1989; 004059.
 PR 10-JAN-1989; JP-004059.
 PA (SANW) SANWA KAGAKU KENKYUSHO.
 DR WPI: 90-265445/35.
 DR N-PSDB: Q05764.
 DR cDNA clone of non-A and non-B hepatitis virus (NANB) - contains
 PT CA 53KB nucleotide which codes amino acid sequence of NANB post
 PT transfusion.
 PS Claim 2; Fig 3; 6pp; Japanese.
 CC Antigenic gene product may be produced from an M13mpl8 transformed
 CC expression system and used for vaccination and diagnosis and therapy.
 SQ Sequence 82 AA;

Query Match 61.4%; Score 51; DB 1; Length 82;
 Best Local Similarity 71.4%; Pred. No. 2.14e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 42 kvflfsf 48
 ||:||||
 QY 62 KVLDFDF 68

RESULT 11
 ID R63760 standard; Protein; 280 AA.
 AC R63760;
 DT 15-MAY-1995 (first entry)
 DE sacs anti-terminator.

KW sacs anti-terminator; sacR; sacY; sucrose inducible expression system;
KW Bacillus subtilis; positive regulator gene.
OS Bacillus subtilis.
FH Key Location/Qualifiers
FT misc_difference 32 /note= "Corresponding codon codes Lys"
PN US5350690-A.
PD 27-SEP-1994.
PF 19-JUN-1987; 064933.
PR 19-JUN-1987; US-064933.
PR 22-FEB-1991; US-659217.
PR 28-APR-1992; US-876728.
PA (AMGE-) AMGEN INC.
PI Zukowski MW.
DR WPI: 94-310588/38.
DR N-PSDB; Q74442.
PT Plasmid vector for expressing genes encoding a desired polypeptide - in a Bacillus subtilis host microorganism.
PS Disclosure; Columns 3-4; 18pp; English.
CC Q74442 is the sacY gene which codes for R63760, a transcription anti-terminator. In the presence of sucrose, R63760 renders sacR the sacs transcription terminator non-functional. Q74442 was used in the construction of a plasmid which transformed B. subtilis into a sucrose inducible expression system, for the production of desired polypeptides.
CC desired polypeptides.
SQ Sequence 280 AA;

Query Match 61.4%; Score 51; DB 12; Length 280;
Best Local Similarity 66.7%; Pred. No. 2.14e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 252 keygfefpe 260
QY 62 KVLFDPE 70

RESULT 12
ID W72745 standard; protein; 411 AA.
AC W72745;
DT 11-JAN-1999 (first entry)
DE Drosophila kinesin N-terminal 411 amino acid residues.
KW Drosophila; kinesin; separation; hybridisation; target site;
KW complex mixture; motor protein; actively transported; separated;
KW microtubule.
OS Drosophila sp.
PN US5830659-A.
PD 03-NOV-1998.
PF 13-SEP-1996; 713815.
PR 13-SEP-1996; US-713815.
PA (UTAH) UNIV UTAH RES FOUND.
PI Stewart RJ;
DR WPI: 98-609236/51.
PT Separation of selected molecules, e.g. DNA, from complex mixtures - uses specific apparatus to allow the selected molecule to bind to motor proteins, and be actively transported and separated away along microtubules
PS Claim 3; Column 25-28; 24pp; English.
CC A method has been developed of separating a selected molecule from a mixture of molecules. The method comprises: (a) a separation device comprising a loading reservoir and a receiving reservoir coupled by a channel with microtubules immobilised on its surface and aligned parallel to a longitudinal axis of the channel; (b) loading the loading reservoir with an aqueous solution of the mixture of molecules; (c) adding a motor-ligand composition and ATP to the solution, where the motor-ligand comprises, (i) a processive motor capable of attaching to the immobilised microtubules, and moving in the presence of ATP as source of chemical energy, and (ii) a ligand coupled to the motor protein, where the ligand is capable of binding the selected molecule, so that the ligand binds the selected molecule and the motor protein attaches to the immobilised microtubules and transports the bound selected molecules along the receiving reservoir; and (d) removing the selected molecule from the receiving reservoir. The method and the system are used for the separation of specific molecules from complex mixtures. The molecule to

CC be separated (e.g. DNA) binds to the motor protein due to the presence of the specific binding ligands. Activation of these enables them to travel down a preformed channel in a specially made piece of apparatus. They can then be removed easily without contaminants of other mixture particles.
CC The present sequence represents the N-terminal 411 amino acid residues of Drosophila kinesin for use in the method of the invention.
SQ Sequence 411 AA;

Query Match 61.4%; Score 51; DB 36; Length 411;
Best Local Similarity 100.0%; Pred. No. 2.14e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 51 kvylfd 56
QY 62 KVLFD 67

RESULT 13
ID P90549 standard; protein; 418 AA.
AC P90549;
DT 14-FEB-1990 (first entry)
DE Serotonin 5HT1c receptor.
KW Serotonin; 5-hydroxytryptamine (5HT); 5HT1c receptor.
OS Rattus rattus.
FH Key Location/Qualifiers
FT domain 53..79
FT /label=I
FT /note="transmembrane domain"
FT domain 91..111
FT /label=II
FT /note="transmembrane domain"
FT domain 129..151
FT /label=III
FT /note="transmembrane domain"
FT domain 172..195
FT /label=IV
FT /note="transmembrane domain"
FT domain 215..236
FT /label=V
FT /note="transmembrane domain"
FT domain 276..298
FT /label=VI
FT /note="transmembrane domain"
FT domain 313..335
FT /label=VII
FT /note="transmembrane domain"
FT domain 247
FT /note="39 bp cytoplasmic loop"
PN WO9918149-A.
PD 08-SEP-1989.
PR 28-FEB-1989; U00808.
PR 29-FEB-1988; US-162654.
PA (UYCO-) Columbia University.
PI Axel R Jessell TM;
DR WPI: 89-278308/38.
DR N-PSDB; N90955.
PT DNA encoding serotonin 5HT1c receptor - used for producing protein, antibodies and probes for studying receptor binding and screening drugs.
PS Claim 3; fig. 1; 84pp; English.
CC Sequence codes for rat serotonin 5HT1c receptor. It was sequenced from the cDNA clone PSR-1c.
CC See also P90549-P90554, P92111 and N90955.
SQ Sequence 418 AA;

Query Match 61.4%; Score 51; DB 1; Length 418;
Best Local Similarity 71.4%; Pred. No. 2.14e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 46 lfqfpgd 52
QY 65 LFDPE 71

RESULT 14
ID W72744 standard; protein; 441 AA.
AC W72744;
DT 11-JAN-1999 (first entry)
DE Drosophila kinesin N-terminal 441 amino acid residues.
KW Drosophila: kinesin; separation; hybridisation; target site;
KW complex mixture; motor protein; actively transported; separated;
KW microtubule.
OS Drosophila sp.
PN US5830659-A.
PD 03-NOV-1998.
PF 13-SEP-1996; 713815
PR 13-SEP-1996; US-713815.
PA (UTAH) UNIV UTAH RES FOUND.
PI Stewart RJ;
DR WPI; 98-609236/51.
PT Separation of selected molecules, e.g. DNA, from complex mixtures -
PT uses specific apparatus to allow the selected molecule to bind to
PT motor proteins, and be actively transported and separated away along
PT micro:tubules
PS Claim 3; Column 23-26; 24pp; English.
CC A method has been developed of separating a selected molecule from a
CC mixture of molecules. The method comprises: (a) a separation device
CC comprising a loading reservoir and a receiving reservoir coupled by a
CC channel with microtubules immobilised on its surface and aligned parallel
CC to a longitudinal axis of the channel; (b) loading the loading reservoir
CC with an aqueous solution of the mixture of molecules; (c) adding a motor-
CC ligand composition and ATP to the solution, where the motor-ligand
CC comprises: (i) a processive motor capable of attaching to the immobilised
CC microtubules, and moving in the presence of ATP as source of chemical
CC energy, and (ii) a ligand coupled to the motor protein, where the ligand
CC is capable of binding the selected molecule, so that the ligand binds the
CC selected molecule and the motor protein attaches to the immobilised
CC microtubules and transports the bound selected molecules along the
CC receiving reservoir; and (d) removing the selected molecule from the
CC receiving reservoir. The method and the system are used for the
CC separation of specific molecules from complex mixtures. The molecule to
CC be separated (e.g. DNA) binds to the motor protein due to the presence of
CC the specific binding ligands. Activation of these enables them to travel
CC down a preformed channel in a specially made piece of apparatus. They can
CC then be removed easily without contaminants of other mixture particles.
CC The present sequence represents the N-terminal 441 amino acid residues of
CC Drosophila kinesin for use in the method of the invention.
SQ Sequence 441 AA;
Query Match 61.4%; Score 51; DB 36; Length 441;
Best Local Similarity 100.0%; Pred. No. 2.14e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 51 kvylfd 56
Qy 62 kvylfd 67
RESULT 15
ID R60697 standard; Protein; 453 AA.
AC R60697;
DT 25-JUN-1995 (first entry)
DE Fragment of the rat serotonin receptor.
KW 5-HT; expression.
OS Rattus rattus.
PN WO9421789-A.
PD 29-SEP-1994.
PF 28-FEB-1994; U02388.
PR 25-MAR-1993; US-038662.
PA (REGC) UNIV CALIFORNIA.
FI Betlach MC, Turner GJ;
DR WPI; 94-317010/39.
DR N-PSDB; Q73589.
PT Expression of heterologous proteins in halo-bacteria - using
PT regulatory and stop sequences from halo-bacteria, pref. the
PT bacterio-rhodopsin gene.
PS Disclosure: Fig 10; 118pp; English.

CC The sequence is that of the rat serotonin receptor fragment.
CC This is used to exemplify a new expression vector for producing
CC heterologous polypeptides in a halobacterial host.
CC See also R60691-9.
SQ Sequence 453 AA;

Query Match 61.4%; Score 51; DB 12; Length 453;
Best Local Similarity 71.4%; Pred. No. 2.14e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 39 lfqfpg 45
Qy 65 lfdfpeg 71

Search completed: Thu Jul 8 18:14:30 1999
Job time : 18 secs.

WQSEAL

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:13:40 1999; MasPar time 4.93 Seconds
Tabular output not generated. 81.323 Million cell updates/sec

Title: >US-09-041-236-2
Description: (62-71) from US09041236.pep (5 of 45)
Perfect Score: 83
Sequence: 1 KVLDFDFPEG 10

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 26.048; Variance 40.214; scale 0.648

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	66	79.5	775	2	E70320 polyribonucleotide nu	2.74e-01
2	64	77.1	561	2	JC2436 5'-nucleotidase (EC 3	6.24e-01
3	62	74.7	560	2	H70466 serine proteinase (EC	1.40e+00
4	60	72.3	213	2	S74247 CPDdiacylglycerol--in	3.10e+00
5	58	69.9	88	2	B42505 K3L protein - vaccini	6.76e+00
6	58	69.9	1172	2	C70619 probable lysX protein	6.76e+00
7	57	68.7	88	1	WMVZK2 K2 protein - vaccinia	9.92e+00
8	56	67.5	374	2	E70346 mannosyltransferase B	1.45e-01
9	56	67.5	1201	2	T00444 hypothetical protein	1.45e-01
10	55	66.3	349	1	WNADEF3 early E1A 27K protein	2.10e+01
11	55	66.3	282	2	A70078 conserved hypothetical	2.10e+01
12	55	66.3	446	2	G64227 glycine--tRNA ligase	2.10e+01
13	55	66.3	653	2	T03102 semaphorin homolog A3	2.10e+01
14	55	66.3	1207	2	C70013 conserved hypothetical	2.10e+01
15	55	66.3	2523	2	F70846 probable pPE protein	2.10e+01
16	54	65.1	514	1	DWEBTT threonine dehydratase	3.04e+01
17	54	65.1	514	1	DNECTS threonine dehydratase	3.04e+01
18	54	65.1	516	2	S44694 hypothetical protein	3.04e+01
19	54	65.1	2295	2	B17621 probable membrane ass	3.04e+01
20	53	63.9	230	2	S22406 hydrogenase (EC 1.18	4.38e-01
21	53	63.9	449	2	S73808 glycine--tRNA ligase	4.38e-01
22	53	63.9	518	2	S75811 gamma-glutamyltransfe	4.38e-01
23	53	63.9	633	2	G71241 probable alpha-amylas	4.38e-01

24	53	63.9	1068	2	S73091 hypothetical protein	4.38e+01
25	52	62.7	153	2	D42513 JIR protein - vaccini	6.28e+01
26	52	62.7	159	2	S33092 JIR protein - variola	6.28e+01
27	52	62.7	173	2	F70481 hypothetical protein	6.28e+01
28	52	62.7	193	2	S64290 hypothetical protein	6.28e+01
29	52	62.7	250	2	G70968 probable methyltransf	6.28e+01
30	52	62.7	264	2	D69497 ABC transporter, perm	6.28e+01
31	52	62.7	357	2	S72575 hypothetical protein	6.28e+01
32	52	62.7	415	3	A40727 rad4+/cut5+ product -	6.28e+01
33	52	62.7	439	2	A71873 adenosylmethionine-8-	6.28e+01
34	52	62.7	579	2	S22584 rad4 protein - fissio	6.28e+01
35	52	62.7	682	2	S50436 hypothetical COA	6.28e+01
36	52	62.7	786	2	S71091 acetyl-CoA carboxylas	6.28e+01
37	52	62.7	1236	1	VHWVME structural polyprotei	6.28e+01
38	52	62.7	1239	1	VHWVEE structural polyprotei	6.28e+01
39	52	62.7	1240	1	VHWVEE structural polyprotei	6.28e+01
40	52	62.7	1241	2	S26373 genome polyprotein -	6.28e+01
41	52	62.7	1242	2	A56605 structural polyprotei	6.28e+01
42	52	62.7	1245	1	VHWVB structural polyprotei	6.28e+01
43	52	62.7	1245	1	VHWVB2 structural polyprotei	6.28e+01
44	52	62.7	1245	1	VHWVB2 structural polyprotei	6.28e+01
45	52	62.7	2339	2	S41121 acetyl-CoA carboxylas	6.28e+01

ALIGNMENTS

RESULT 1
ENTRY E70320 #type complete
TITLE polyribonucleotide nucleotidyltransferase - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 04-Sep-1998
ACCESSIONS E70320
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
#cross-references MUID:98196666
#accession E70320
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-775 #label AQF
#cross-references GB:AE000679; NID:g2982936; PID:g2982946; GB:AE000657
#experimental_source strain VF5
GENETICS
#gene pfpA
CLASSIFICATION #superfamily polyribonucleotide nucleotidyltransferase alpha chain
SUMMARY #length 775 #molecular-weight 86488 #checksum 4882
Query Match 79.5%; Score 66; DB 2; Length 775;
Best Local Similarity 80.0%; Pred. No. 2.74e-01;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 595 KVLDFDFPEG 604
Qy 62 KVLDFDFPEG 71
RESULT 2
ENTRY JC2436 #type complete
TITLE 5'-nucleotidase (EC 3.1.3.5), cytosolic - human
ALTERNATE_NAMES cytosolic purine 5'-nucleotidase
ORGANISM #formal_name Homo sapiens #common_name man
DATE 15-Feb-1995 #sequence_revision 05-Apr-1995 #text_change 24-Sep-1998
ACCESSIONS JC2436
REFERENCE JC2436

```
#authors Oka, J.; Matsumoto, A.; Hosokawa, Y.; Inoue, S.
#journal Biochem. Biophys. Res. Commun. (1994) 205:917-922
#title Molecular cloning of human cytosolic purine 5'-nucleotidase.
#cross-references MUID:95091838
#accession JC2436
##molecule_type mRNA
##residues 1-561 ##label OKA
##cross-references DBJ:D38524; NID:g633070; PID:dl008110; PID:g633071
COMMENT This enzyme is a soluble nucleotidase. It hydrolyzes IMP and other
purine nucleotides and is allosterically activated by various
compounds including ATP.
COMMENT This enzyme has a critical role in the maintenance of a constant
composition of intracellular purine/pyrimidine nucleotides in
cooperation with other nucleotidases.
GENETICS
#gene GDB:NT5CP
##cross-references GDB:569367
KEYWORDS cytosol; phosphoric monoester hydrolase
SUMMARY #length 561 #molecular-weight 64969 #checksum 6469
Query Match 77.1%; Score 64; DB 2; Length 561;
Best Local Similarity 70.0%; Pred. No. 6.24e-01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 260 MTLFDFPHG 269
:|||||
QY 62 KVLDFDFPEG 71
RESULT 3
ENTRY H70466 #type complete
TITLE serine proteinase (EC 3.4.21.-) - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
22-May-1998
ACCESSIONS H70466
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
Keller, M.; Aulaj, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MUID:98196666
#accession H70466
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-560 ##label AOF
##cross-references GB:AE000764; NID:g2984190; PID:g2984198; GB:AE000657
#experimental_source strain VF5
GENETICS
#gene aprV
KEYWORDS hydrolase; serine proteinase
SUMMARY #length 560 #molecular-weight 61463 #checksum 1470
Query Match 74.7%; Score 62; DB 2; Length 560;
Best Local Similarity 70.0%; Pred. No. 1.40e+00;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 39 KVLDDLDPDG 48
:|||||
QY 62 KVLDFDFPEG 71
RESULT 4
ENTRY S74247 #type complete
TITLE CDPdiacylglycerol--inositol 3-phosphatidyltransferase (EC
2.7.8.11) - rat
ALTERNATE_NAMES phosphatidylinositol synthase
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change
```

```
04-Sep-1998
ACCESSIONS S74247
REFERENCE S74247
#authors Tanaka, S.; Nikawa, J.; Imai, H.; Yamashita, S.; Hosaka, K.
#journal FEBS Lett. (1996) 393:89-92
#title Molecular cloning of rat phosphatidylinositol synthase cDNA
by functional complementation of the yeast Saccharomyces
cerevisiae p15 mutation.
#cross-references MUID:96397537
#accession S74247
##molecule_type mRNA
##residues 1-213 ##label TAN
##cross-references EMBL:D82928; NID:g1620878; PID:dl012301; PID:g1620879
#experimental_source strain Wistar; tissue type brain
CLASSIFICATION #superfamily CDPdiacylglycerol--inositol
3-phosphatidyltransferase
KEYWORDS membrane protein; transferase
SUMMARY #length 213 #molecular-weight 23613 #checksum 3429
Query Match 72.3%; Score 60; DB 2; Length 213;
Best Local Similarity 66.7%; Pred. No. 3.10e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 155 LXLNFSEEG 163
:|||||
QY 63 VYLDFDFPEG 71
RESULT 5
ENTRY B42505 #type complete
TITLE K3L protein - vaccinia virus (strain Copenhagen)
ORGANISM #formal_name vaccinia virus
#note host Homo sapiens (man)
DATE 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change
08-Apr-1994
ACCESSIONS B42505
REFERENCE A33172
#authors Johnson, G.P.
#submission submitted to GenBank, June 1990
#accession B42505
##status preliminary
##molecule_type DNA
##residues 1-88 ##label JOH
CLASSIFICATION #superfamily vaccinia virus K2 protein
SUMMARY #length 88 #molecular-weight 10496 #checksum 5796
Query Match 69.9%; Score 58; DB 2; Length 88;
Best Local Similarity 55.8%; Pred. No. 6.76e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 28 IYLFDFPHS 36
:|||||
QY 63 VYLDFDFPEG 71
RESULT 6
ENTRY C70619 #type complete
TITLE probable lysX protein - Mycobacterium tuberculosis (strain
H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS C70619
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; K.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
```

```
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession C70619
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-1172 ##label COL
##cross-references GB:285982; GB:AL123456; NID:g3261718; PID:e300638;
PID:g1838991
##experimental_source strain H37Rv
GENETICS
#gene lysX
#summary #length 1172 #molecular-weight 128238 #checksum 5958
Query Match 69.9%; Score 58; DB 2; Length 1172;
Best Local Similarity 71.4%; Pred. No. 6.76e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 118 YLNFDPD 124
QY 64 YLFDPE 70
RESULT 7
ENTRY MWVK2 #type complete
TITLE K2 protein - vaccinia virus (strain WR)
ORGANISM #formal_name vaccinia virus
DATE 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change
20-Mar-1998
ACCESSION JS0212
REFERENCE JS0211
#authors Boursnell, M.E.G.; Foulds, I.J.; Campbell, J.I.; Binns, M.M.
#journal J. Gen. Virol. (1988) 69:2995-3003
#title Non-essential genes in the vaccinia virus HindIII K fragment:
a gene related to serine protease inhibitors and a gene
related to the 37K vaccinia virus major envelope antigen.
#cross-references MUID:89067908
#accession JS0212
##molecule_type DNA
##residues 1-88 ##label BOU
##cross-references GB:D00382; NID:g222704; PID:d1000739; PID:g222707
CLASSIFICATION #superfamily vaccinia virus K2 protein
SUMMARY #length 88 #molecular-weight 10556 #checksum 5328
Query Match 68.7%; Score 57; DB 1; Length 88;
Best Local Similarity 71.4%; Pred. No. 9.92e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 28 IYLFDP 34
QY 63 YLFDPE 69
RESULT 8
ENTRY E70346 #type complete
TITLE mannosyltransferase B - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
08-May-1998
ACCESSION E70346
REFERENCE A70300
#authors Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, W.G.;
Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MUID:98196666
#accession E70346
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-374 ##label AQP
##cross-references GB:AE000693; NID:g2983148; PID:g2983150; GB:AE000657
##experimental_source strain VF5
GENETICS
#gene mtfB
#summary #length 374 #molecular-weight 43869 #checksum 9875
Query Match 67.5%; Score 56; DB 2; Length 374;
Best Local Similarity 87.5%; Pred. No. 1.45e+01;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 93 KYLFFFP 100
QY 62 KYLFDFF 69
RESULT 9
ENTRY T00444 #type complete
TITLE hypothetical protein T14N5.1 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
01-Feb-1999
ACCESSION T00444
REFERENCE T14152
#authors Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.;
Conway, A.R.; Au, M.; Araujo, R.; Buehler, E.; Dewar, K.;
Feng, J.; Kim, C.; Li, Y.; Oji, O.; Osborne, B.I.; Shinn,
P.; Sun, H.; Toriumi, M.; Vysotskaia, V.S.; Yu, G.; Ecker,
J.; Theologis, A.; Davis, R.W.
#submission submitted to the EMBL Data Library, September 1998
#accession T00444
##molecule_type DNA
##residues 1-1201 ##label FED
##cross-references EMBL:AC004260; NID:g3176694; PID:g3540195
GENETICS
#map_position I
#introns 251/1; 569/1; 601/2; 858/1
#note T14N5.1
SUMMARY #length 1201 #molecular-weight 139308 #checksum 3109
Query Match 67.5%; Score 56; DB 2; Length 1201;
Best Local Similarity 71.4%; Pred. No. 1.45e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 973 FLFFPE 979
QY 64 YLFDPE 70
RESULT 10
ENTRY WMAFP3 #type complete
TITLE early E1A 27K protein - human adenovirus 40
CONTAINS early E1A 15K protein; early E1A 24K protein
ORGANISM #formal_name Mastadenovirus h40 #common_name human adenovirus
40
#note host Homo sapiens (man)
DATE 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
22-Jan-1999
ACCESSION A27333; A29195; B27333; C27333
REFERENCE A27333
#authors van Loon, A.E.; Ligtenberg, M.; Reemst, A.M.C.B.; Sussenbach,
J.S.; Roziijn, T.H.
#journal Gene (1987) 58:109-126
#title Structure and organization of the left-terminal DNA regions
of fastidious adenovirus types 40 and 41.
#cross-references MUID:88084437
#accession A27333
##molecule_type DNA
```

```

##residues 1-249 ##label VAN
REFERENCE A94383
#authors Ishino, M.; Ohashi, Y.; Emoto, T.; Sawada, Y.; Fujinaga, K.
#journal Virology (1988) 165:95-102
#title Characterization of adenovirus type 40 El region.
#cross-references MUID:88265890
#accession A29195
##molecule_type DNA
##residues 1-249 ##label ISH
##cross-references GB:L19443; NID:G303969; PID:G303971
CLASSIFICATION #superfamily adenovirus early E1A protein
KEYWORDS alternative splicing; DNA binding; early protein;
transcription regulation; zinc finger
FEATURE
1-148,177-249 #product early E1A 24K protein #status predicted #label
EA2\
1-62,177-249 #product early E1A 15K protein #status predicted #label
EA1\
SUMMARY #length 249 #molecular-weight 27481 #checksum 2202
Query Match 66.3%; Score 55; DB 1; Length 249;
Best Local Similarity 85.7%; Pred. No. 2.10e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 24 YVDFPE 30
I:|||||
QY 64 YLDFPE 70

RESULT 11
ENTRY #type complete
TITLE conserved hypothetical protein yx1m - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
ACCESSIONS A70078
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Allonli, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Hatech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mazel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Skowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Nature (1997) 390:249-256
#journal

```

```

#title The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession A70078
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-382 ##label KUN
##cross-references GB:299124; GB:AL009126; NID:G2636442; PID:el184637;
PID:G2636458
##experimental_source strain 168
GENETICS
#gene yx1m
CLASSIFICATION #superfamily hypothetical protein yx1m
SUMMARY #length 382 #molecular-weight 41836 #checksum 6590
Query Match 66.3%; Score 55; DB 2; Length 382;
Best Local Similarity 70.0%; Pred. No. 2.10e+01;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Db 29 KVQDFDGG 38
I:|||||
QY 62 KVLDFDPE 71

RESULT 12
ENTRY #type complete
TITLE glycine--tRNA ligase (EC 6.1.1.14) - Mycoplasma genitalium
ALTERNATE_NAMES glycyi-tRNA synthetase
ORGANISM #formal_name Mycoplasma genitalium
DATE 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
18-Sep-1998
ACCESSIONS G64227
REFERENCE A64200
#authors Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
C.A.; Venter, J.C.
#journal Science (1995) 270:397-403
#title The minimal gene complement of Mycoplasma genitalium.
#cross-references MUID:96026346
#accession G64227
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-446 ##label TIGR
##cross-references GB:U39703; GB:L43967; NID:G1045933; PID:G1045942;
TIGR:MG351
##experimental_source strain G-37
GENETICS
#genetic_code SGC3
CLASSIFICATION #superfamily Mycoplasma genitalium glycine--tRNA ligase
KEYWORDS aminoacyl-tRNA synthetase; ligase; protein biosynthesis
SUMMARY #length 446 #molecular-weight 52295 #checksum 795
Query Match 66.3%; Score 55; DB 2; Length 446;
Best Local Similarity 62.5%; Pred. No. 2.10e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 265 FLNFPFG 272
I:|||||
QY 64 YLDFDPE 71

RESULT 13
ENTRY #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1

```

```
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
ACCESSIONS T03102
REFERENCE Z14840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-653 ##label ENS
##cross-references EMBL:AF003370; NID:g2337967; PID:g2337970
SUMMARY #length 653 #molecular-weight 73645 #checksum 5501

Query Match 66.3%; Score 55; DB 2; Length 653;
Best Local Similarity 55.6%; Pred. No. 2.10e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 117 IYLFDFAHs 125
QY 63 VYLFDFPEG 71

RESULT 14
ENTRY #type complete
TITLE conserved hypothetical protein yuka - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
C70013
A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Aloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denicot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entlan, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moesti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Tarpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession C70013
##status preliminary; nucleic acid sequence not shown;
translation not shown
```

```
##molecule_type DNA
##residues 1-1207 ##label KUN
##cross-references GB:Z99120; GB:AL009126; NID:g2635613; PID:e1184264;
##experimental_source strain 168
GENETICS
#gene yuka
SUMMARY #length 1207 #molecular-weight 139251 #checksum 2184

Query Match 66.3%; Score 55; DB 2; Length 1207;
Best Local Similarity 66.7%; Pred. No. 2.10e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 757 VYIFDFGNG 765
QY 63 VYLFDFPEG 71

RESULT 15
ENTRY #type complete
TITLE probable pPE protein - Mycobacterium tuberculosis (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
A70846
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekaiia, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsbey, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellon, S.; Squares, S.; Szares, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrall, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession F70846
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-2523 ##label COL
##cross-references GB:AL021841; GB:AL123456; NID:g3261517; PID:e1251175;
##experimental_source strain H37Rv
GENETICS
#gene PPE
SUMMARY #length 2523 #molecular-weight 250344 #checksum 6023

Query Match 66.3%; Score 55; DB 2; Length 2523;
Best Local Similarity 55.6%; Pred. No. 2.10e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1218 LFLFDFRDG 1226
QY 63 VYLFDFPEG 71

Search completed: Thu Jul 8 18:13:51 1999
Job time : 11 secs.
```

THIS PAGE BLANK (USPTO)

(TN)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:12:39 1999; MasPar time 3.35 Seconds
Tabular output not generated. 84.327 Million cell updates/sec

Title: >US-09-041-236-2
Description: (62-71) from US09041236.pep (5 of 45)
Perfect Score: 83
Sequence: 1 KVLDFDFPEG 10

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 26.674; Variance 35.920; scale 0.743

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	64	77.1	561	1	5NTC_HUMAN CYTOSOLIC PURINE 5'-NU	1.46e+01
2	61	73.5	128	1	YF9A_METUA HYPOTHETICAL PROTEIN M	5.64e+01
3	58	69.9	88	1	VK03_VACCC PROTEIN K3	2.10e+00
4	57	68.7	88	1	VK03_VACCV PROTEIN K2	3.22e+00
5	57	68.7	213	1	PIS_HUMAN CDP-DIACYLGLYCEROL--IN	3.22e+00
6	56	67.5	66	1	YK89_CAEEL HYPOTHETICAL 8.0 KD PR	4.91e+00
7	55	66.3	249	1	EIA_ADE40 EARLY EIA 27 KD PROTEI	7.44e+00
8	55	66.3	382	1	YXIM_BACSU HYPOTHETICAL 41.8 KD P	7.44e+00
9	55	66.3	446	1	SYG_MYCGE GLYCYL-TRNA SYNTHETASE	7.44e+00
10	55	66.3	513	1	YH01_HAEIN THREONINE DEHYDRATASE	7.44e+00
11	54	65.1	194	1	YH01_RHISN THREONINE DEHYDRATASE	1.12e+01
12	54	65.1	514	1	THD1_SALTY THREONINE DEHYDRATASE	1.12e+01
13	54	65.1	514	1	THD1_ECOLI THREONINE DEHYDRATASE	1.12e+01
14	54	65.1	516	1	YVNF_AZOCH HYPOTHETICAL PROTEIN I	1.12e+01
15	54	65.1	550	1	THDH_ARXAD THREONINE DEHYDRATASE	1.12e+01
16	53	63.9	230	1	CYBH_WOLSU QUINONE-REACTIVE NI/FE	1.68e+01
17	53	63.9	449	1	SYG_MYCPN GLYCYL-TRNA SYNTHETASE	1.68e+01
18	53	63.9	518	1	GGT_STNY3 GAMMA-GLUTAMYLTRANSFERP	1.68e+01
19	53	63.9	532	1	P72_MYCMY IMMUNODOMINANT PROTEIN	1.68e+01
20	52	62.7	153	1	VJ01_VACCV PROTEIN J1 (PROTEIN F7	2.51e+01
21	52	62.7	153	1	VJ01_VACCC PROTEIN J1.	2.51e+01
22	52	62.7	159	1	VJ01_VARV PROTEIN J1.	2.51e+01
23	52	62.7	193	1	YGL1_YEAST HYPOTHETICAL 22.2 KD P	2.51e+01

24	52	62.7	436	1	BIOA_HELPV ADENOSYLMETHIONINE-8-A	2.51e+01
25	52	62.7	607	1	VEL_HPV23 REPLICATION PROTEIN E1	2.51e+01
26	52	62.7	608	1	VEL_HPV22 REPLICATION PROTEIN E1	2.51e+01
27	52	62.7	621	1	Y241_MYCPN HYPOTHETICAL PROTEIN M	2.51e+01
28	52	62.7	648	1	RAD4_SCHPO RAD4 PROTEIN (CUT5 PRO	2.51e+01
29	52	62.7	682	1	YEC3_YEAST HYPOTHETICAL 78.3 KD P	2.51e+01
30	52	62.7	714	1	YFE7_YEAST HYPOTHETICAL 82.2 KD P	2.51e+01
31	52	62.7	1188	1	YEC5_YEAST HYPOTHETICAL 136.1 KD	2.51e+01
32	52	62.7	1236	1	POLS_WEEV STRUCTURAL POLYPROTEIN	2.51e+01
33	52	62.7	1239	1	POLS_EEEV STRUCTURAL POLYPROTEIN	2.51e+01
34	52	62.7	1240	1	POLS_EEEV3 STRUCTURAL POLYPROTEIN	2.51e+01
35	52	62.7	1245	1	POLS_SINDO STRUCTURAL POLYPROTEIN	2.51e+01
36	52	62.7	1245	1	POLS_SINDO STRUCTURAL POLYPROTEIN	2.51e+01
37	52	62.7	2483	1	COA2_HUMAN ACETYL-COA CARBOXYLASE	2.51e+01
38	51	61.4	396	1	YJRL_YEAST HYPOTHETICAL 43.0 KD P	3.72e+01
39	51	61.4	459	1	SH2C_MOUSE 5-HYDROXYTRYPTAMINE 2C	3.72e+01
40	51	61.4	798	1	VPL6_YEAST VACUOLAR PROTEIN SORTI	3.72e+01
41	51	61.4	831	1	VPL1_HUMAN CLATHRIN-COATED VESICL	3.72e+01
42	51	61.4	975	1	KINH_DROME KINESIN HEAVY CHAIN.	3.72e+01
43	51	61.4	1023	1	GLT_DROME GLUTACTIN PRECURSOR.	3.72e+01
44	51	61.4	1178	1	ITAZ_MOUSE PLATELET MEMBRANE GLYC	3.72e+01
45	51	61.4	2233	1	COAC_YEAST ACETYL-COA CARBOXYLASE	3.72e+01

ALIGNMENTS

RESULT 1
ID 5NTC_HUMAN STANDARD; PRT; 561 AA.
AC P49902;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CYTOSOLIC PURINE 5'-NUCLEOTIDASE (EC 3.1.3.5).
GN NT5CP.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE; 95091838.
RA OKA J., MATSUMOTO A., HOSOKAWA Y., INOUE S.;
RT "Molecular cloning of human cytosolic purine 5'-nucleotidase.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 205:917-922(1994).
CC -!- FUNCTION: MAY HAVE A CRITICAL ROLE IN THE MAINTENANCE OF A
CC CONSTANT COMPOSITION OF INTRACELLULAR PURINE/PYRIMIDINE
CC NUCLEOTIDES IN COOPERATION WITH OTHER NUCLEOTIDASES.
CC PREFERENTIALLY HYDROLYZES INOSINE 5-PRIME-MONOPHOSPHATE (IMP) AND
CC OTHER PURINE NUCLEOTIDES.
CC -!- CATALYTIC ACTIVITY: A 5'-RIBONUCLEOTIDE + H(2)O -> A
CC RIBONUCLEOSIDE + ORTHOPHOSPHATE.
CC -!- ENZYME REGULATION: ALLOSTERICALLY ACTIVATED BY VARIOUS COMPOUNDS,
CC INCLUDING ATP.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- DISEASE: ABERRATION OF THIS ENZYME MAY BE RELATED TO THE URATE
CC PRODUCTION IN HYPERURICEMIA AND GOUT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: D38524; G633071; -
CC MM: 500417; -
CC
CC HYDROLASE; ALLOSTERIC ENZYME.
KW DOMAIN 549 561 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 561 AA; 64969 MW; 88673837 CRC32;

Query Match 77.1%; Score 64; DB 1; Length 561;
Best Local Similarity 70.0%; Pred. No. 1.46e+01;

```
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 260 MTLFDFPHG 269
   : |||||
QY 62 KVLFDPEP 71

RESULT 2
ID YF9A-METJA STANDARD; PRT; 128 AA.
AC P81325;
AT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE 13-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN MJ1597.1.
GN MJ1597.1.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEA: EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
OC METHANOCOCCUS.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE: 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
RA SCOTT J.L., GEOCHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT Jannaschii."
RL SCIENCE 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67600; G2826438; -.
DR TIGR: MJ1597.1; -.
DR KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 128 AA; 15255 MW; 16613DD5 CRC32;

Query Match 73.5%; Score 61; DB 1; Length 128;
Best Local Similarity 75.0%; Pred. No. 5.64e-01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 43 RVLFSFP 50
   : |||||
QY 62 KVLFDPEP 69

RESULT 3
ID VK03_VACCC STANDARD; PRT; 88 AA.
AC P20639;
AT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)
DE PROTEIN K3.
GN K3L.
OS VACCINIA VIRUS (STRAIN COPENHAGEN).
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 91021027.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RT "The complete DNA sequence of vaccinia virus."
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
VIROLOGY 179:247-266(1990).
[2]
RN COMPLETE GENOME.
RP GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RA VIROLOGY 179:517-563(1990).
[3]
RN RESISTANCE TO INTERFERON.
RP MEDLINE: 91272507.
RX BEATTIE E., TARTAGLIA J., PAOLETTI E.;
RT "Vaccinia virus-encoded eIF-2 alpha homolog abrogates the antiviral
RT effect of interferon."
RL VIROLOGY 183:419-422(1991).
CC -!- DISEASE: PLAYS AN INTEGRAL ROLE IN THE RESISTANCE OF THE VIRUS
CC TO INTERFERON. MAY BIND COMPETITIVELY TO THE P1 KINASE TO BLOCK
CC IF2-ALPHA PHOSPHORYLATION.
CC -!- SIMILARITY: TO N-TERMINAL PART OF EUKARYOTIC TRANSLATION
CC INITIATION FACTOR 2 ALPHA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M35027; G335357; -.
DR PIR: B42505; B42505.
DR PFAM: PF00375; S1; 1.
DR SQ SEQUENCE 88 AA; 10496 MW; 24C9A143 CRC32;

Query Match 69.9%; Score 58; DB 1; Length 88;
Best Local Similarity 55.6%; Pred. No. 2.10e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 28 IYLFDPHS 36
   : |||||
QY 63 VYLFDFPEP 71

RESULT 4
ID VK03_VACCV STANDARD; PRT; 88 AA.
AC P18378;
AT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)
DE PROTEIN K3 (PROTEIN K2).
GN K3L.
OS VACCINIA VIRUS (STRAIN WR).
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 89067908.
RA BOURSNEILL M.E.G., FOULDS I.J., CAMPBELL J.I., BINNS M.M.;
RT "Non-essential genes in the vaccinia virus HindIII K fragment: a gene
RT related to serine protease inhibitors and a gene related to the 37K
RT vaccinia virus major envelope antigen."
RL J. GEN. VIROL. 69:2995-3003(1988).
CC -!- DISEASE: PLAYS AN INTEGRAL ROLE IN THE RESISTANCE OF THE VIRUS
CC TO INTERFERON. MAY BIND COMPETITIVELY TO THE P1 KINASE TO BLOCK
CC IF2-ALPHA PHOSPHORYLATION.
CC -!- SIMILARITY: TO N-TERMINAL PART OF EUKARYOTIC TRANSLATION
CC INITIATION FACTOR 2 ALPHA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC -----
DR EMBL; D00382; G222707;
DR PIR; JS0212; WWZK2.
DR PFAM; PF00575; S1; 1
SQ SEQUENCE 88 AA; 10556 MW; FA4DCB53 CRC32;

Query Match 68.7%; Score 57; DB 1; Length 88;
Best Local Similarity 71.4%; Pred. No. 3.22e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 28 IYLFDP 34
:||||:
QY 63 VYLFDF 69

RESULT 5
ID PIS_HUMAN STANDARD; PRT; 213 AA.
AC O14735;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CDP-DIACYLGLYCEROL--INOSITOL 3-PHOSPHATIDYLTRANSFERASE (EC 2.7.8.11)
DE (PHOSPHATIDYLINOSITOL SYNTHASE).
GN PIS.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98070552.
RA LYKIDIS A., JACKSON P.D., ROCK C.O., JACKOWSKI S.;
RT "The role of CDP-diacylglycerol synthetase and phosphatidylinositol
RT synthase activity levels in the regulation of cellular
RT phosphatidylinositol content.";
RL J. BIOL. CHEM. 272:33402-33409(1997).
CC -!- CATALYTIC ACTIVITY: CDP-DIACYLGLYCEROL + MYO-INOSITOL = CMP +
CC PHOSPHATIDYL-ID-MYO-INOSITOL.
CC -!- SIMILARITY: BELONGS TO THE CDP-ALCOHOL PHOSPHATIDYLTRANSFERASE
CC CLASS-I FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF014807; G2338732;
DR PROSITE; PS00379; CDP-ALCOHOL_P_TRANSF; 1.
DR PFAM; PF01066; CDP-OH_P_TRANSF; 1.
DR TRANSFERASE; PHOSPHOLIPID BIOSYNTHESIS.
SQ SEQUENCE 213 AA; 23538 MW; B93CDC2B CRC32;

Query Match 68.7%; Score 57; DB 1; Length 213;
Best Local Similarity 66.7%; Pred. No. 3.22e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 155 LYLFFHSE 163
:||||:
QY 63 VYLFDFEG 71

RESULT 6
ID YK89_CABEL STANDARD; PRT; 66 AA.
AC P34354;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 8.0 KD PROTEIN C30A5.9 IN CHROMOSOME III.
GN C30A5.9.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIIDA;
```

```
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; LI0990; -; NOT_ANNOTATED_CDS.
DR WORMPEP; C30A5.9; CE00556.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 66 AA; 8040 MW; 9ED32C41 CRC32;

Query Match 67.5%; Score 56; DB 1; Length 66;
Best Local Similarity 50.0%; Pred. No. 4.91e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 15 VFLYERFP 22
:||||:
QY 63 VYLFDFPE 70

RESULT 7
ID ELA_ADE40 STANDARD; PRT; 249 AA.
AC P10341;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE EARLY ELA 27 KD PROTEIN [CONTAINS: EARLY ELA 25 KD PROTEIN].
OS HUMAN ADENOVIRUS TYPE 40.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; ADENOVIRIDAE; MASTADENOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88084437.
RA VAN LOON A.E., LIGTENBERG M., REEMST A.M.C.B., SUSSENBACH J.S.,
RA ROZIJN T.H.;
RT "Structure and organization of the left-terminal DNA regions of
RT fastidious adenovirus types 40 and 41.";
RL GENE 58:109-126(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88265890.
RA ISHINO M., OHASHI Y., EMOTO T., SAWADA Y., FUJINAGA K.;
RT "Characterization of adenovirus type 40 E1 region.";
RL VIROLOGY 165:95-102(1988).
CC -!- FUNCTION: TRANS-ACTIVATES EARLY VIRAL PROMOTERS AND SOME CELLULAR
CC PROMOTERS.
CC -!- ALTERNATIVE PRODUCTS: THESE PROTEINS ARE PRODUCTS OF TWO DIFFERENT
CC MRNAS SPLICED FROM THE SAME PRIMARY TRANSCRIPT FROM THE E1 REGION
CC OF THE GENOME.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: M21276; G209765; -
DR EMBL: M18288; G209771; -
DR EMBL: M18288; G209770; -
DR EMBL: L19443; G303971; -
DR EMBL: L19443; G303972; -
DR PIR: C27333; WMADEF3
KW TRANSCRIPTION REGULATION; ACTIVATOR; EARLY PROTEIN;
KW ALTERNATIVE SPLICING; ZINC-FINGER; DNA-BINDING.
FT ZN_FING 145 165 C4-TYPE.
FT VARSPLIC 149 176 MISSING (IN EARLY ELA 25 KD PROTEIN).
FT CONFLICT 38 38 M -> I (IN REF. 2).
FT CONFLICT 211 211 I -> T (IN REF. 2).
SQ SEQUENCE 249 AA; 27481 MW; 07ALCB22 CRC32;

Query Match 66.3%; Score 55; DB 1; Length 249;
Best Local Similarity 85.7%; Pred. No. 7.44e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 24 YVDFPE 30
|:|||||
Qy 64 YLDFPE 70

RESULT 8

ID YXIM_BACSU STANDARD; PRT; 382 AA.
AC P42304;
DT 01-FEB-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 41.8 KD PROTEIN IN WAPA-LICT INTERGENIC REGION PRECURSOR.
GN YXIM OR SS8D.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / BGSC1A1;
RA YOSHIDA K.-I., SANO H., SEKI S., ODA M., FUJIMURA M., FUJITA Y.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: D83026; D1012359; -
DR EMBL: Z59124; E1184637; -
DR SUBTILIST; BG11142; YXIM.
KW HYPOTHETICAL PROTEIN; SIGNAL.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 382 HYPOTHETICAL PROTEIN YXIM.
SQ SEQUENCE 382 AA; 41836 MW; 4D800BB0 CRC32;

Query Match 66.3%; Score 55; DB 1; Length 382;
Best Local Similarity 70.0%; Pred. No. 7.44e+00;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 29 KYQDFGSG 38
|||||
Qy 62 KYLDFPE 71

RESULT 9

ID SYG_MYCGE STANDARD; PRT; 446 AA.

AC P47493;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE GLYCYL-TRNA SYNTHETASE (EC 6.1.1.14) (GLYCINE--TRNA LIGASE) (GLYRS).
GN GLYS OR MG251.
OS MYCOPLASMA GENITALIUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLIICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE; 96026346.

RA FRASER C.N., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FRIEDLMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
CC SCIENCE 270:397-403(1995).

CC -!- CATALYTIC ACTIVITY: ATP + L-GLYCINE + TRNA(GLY) ~ AMP +
CC PYROPHOSPHATE + L-GLYCYL-TRNA(GLY).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: U39703; G1045942; -
DR TIGR; MG251; -
DR PROSITE; PS00179; AA_TRNA_LIGASE_II_1; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; FALSE_NEG.
DR PFAM; PF00587; trna-synt_2b; 1.
DR HSSP; P56206; LATI
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
SQ SEQUENCE 446 AA; 52295 MW; BD260436 CRC32;

Query Match 66.3%; Score 55; DB 1; Length 446;
Best Local Similarity 62.5%; Pred. No. 7.44e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 265 FLNFPFHG 272
:|||||
Qy 64 YLDFPE 71

RESULT 10

ID THD1_HAEIN STANDARD; PRT; 513 AA.
AC P46493;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE THREONINE DEHYDRATASE BIOSYNTHETIC (EC 4.2.1.16) (THREONINE
DE DEAMINASE).
GN ILVA OR HI0738.1.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,

RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGAN N.S.M.,
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.:
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL SCIENCE 269:496-512(1995).
RN [2]
RA IDENTIFICATION.
RA KOONIN E.V., RUDD K.E.:
RL SUBMITTED (SEP-1995) TO THE SWISS-PROT DATA BANK.
CC -!- FUNCTION: DEAMINATES L-THREONINE, BUT ALSO L-SERINE TO A LESSER
CC EXTENT (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: L-THREONINE + H(2)O = 2-OXOBUTANOATE + NH(3)
CC + H(2)O.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHILE
CC VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME (BY SIMILARITY).
CC -!- PATHWAY: FIRST STEP OF ISOLEUCINE BIOSYNTHESIS.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: TO THREONINE DEHYDRATASE CATABOLIC AND SERINE
CC DEHYDRATASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32757; G1573745; -
CC TIGR: H10738.1; -
CC PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
CC PFAM: PF00291; S_T_dehydratase; 1.
CC PFAM: PF00585; Thr_dehydrat.C; 2.
CC ISOLEUCINE BIOSYNTHESIS; LYASE; PYRIDOXAL PHOSPHATE;
CC ALLOSTERIC ENZYME.
FT BINDING 63 63 PYRIDOXAL PHOSPHATE.
SQ SEQUENCE 513 AA; 56662 MW; 71FF27B6 CRC32;

Query Match 66.3%; Score 55; DB 1: Length 513;
Best Local Similarity 55.6%; Pred. No. 7.44e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 433 RLYTFEPE 441
: : I: I: I:
QY 62 KVLFDPE 70

RESULT 11
ID Y4HQ_RHISN STANDARD; PRT; 194 AA.
AC P50361;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 21.7 KD PROTEIN Y4HQ.
GN Y4HQ.
OS RHIZOBIUM SP. (STRAIN NGR234).
OG PLASMID SYN PNR234A.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; RHIZOBIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97305956.
RA FREIBERG C.A., FELLAY R., BAIRICH A., BROUGHTON W.J., ROSENTHAL A.,
RA PERRET X.:
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL NATURE 387:394-401(1997).
RN [2]
RP SEQUENCE OF 1-52 FROM N.A.
RA ROCHEPEAU P., FELLAY R., BROUGHTON W.J.:

RL SUBMITTED (JUL-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: TO A RHIZOGENES PLASMID PRIAB ORF-3 IN VIRA REGION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X74068; G581537; ALT INIT.
CC EMBL: AE000077; G2182436; -
CC HYPOTHETICAL PROTEIN; PLASMID.
FT CONFLICT 18 18 I -> M (IN REF. 2).
FT CONFLICT 30 30 R -> S (IN REF. 2).
SQ SEQUENCE 194 AA; 21747 MW; C2CDD0EE CRC32;

Query Match 65.1%; Score 54; DB 1: Length 194;
Best Local Similarity 62.5%; Pred. No. 1.12e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 92 YLYDFGDG 99
: : I: I: I:
QY 64 YLDFPEG 71

RESULT 12
ID THDI_SALTY STANDARD; PRT; 514 AA.
AC P20506;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE THREONINE DEHYDRATASE BIOSYNTHETIC (EC 4.2.1.16) (THREONINE
DE DEAMINASE).
GN ILVA.
OS SALMONELLA TYPHIMURIUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88255870.
RA TAILLON B.E., LITTLE R., LAWTHORP R.P.:
RT "Analysis of the functional domains of biosynthetic threonine
RT deaminase by comparison of the amino acid sequences of three
RT wild-type alleles to the amino acid sequence of biodegradative
RT threonine deaminase.";
RL GENE 63:245-252(1988).
RN [2]
RP SEQUENCE OF 1-10 FROM N.A.
RC STRAIN-LT2;
RX MEDLINE: 89326124.
RA LOPES J.M., LAWTHORP R.P.:
RT "Physical identification of an internal promoter, ilvAp, in the
RT distal portion of the ilvGMDA operon.";
RL GENE 76:255-269(1989).
CC -!- CATALYTIC ACTIVITY: L-THREONINE + H(2)O = 2-OXOBUTANOATE + NH(3)
CC + H(2)O.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHILE
CC VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME.
CC -!- PATHWAY: FIRST STEP OF ISOLEUCINE BIOSYNTHESIS.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SIMILARITY: TO THREONINE DEHYDRATASE CATABOLIC AND SERINE
CC DEHYDRATASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

```
CC -----
DR EMBL; M26670; G514966; -.
DR EMBL; M25498; G538506; -.
DR PIR; JT0278; DWEBTT.
DR STYGENE; SG10179; ILVA.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
DR PFAM; PF00291; S_T_dehydratase; 1.
DR PFAM; PF00585; Thr_dehydrat_C; 2.
KW ISOLEUCINE BIOSYNTHESIS; LYASE; PYRIDOXAL PHOSPHATE;
KW ALLOSTERIC ENZYME.
FT BINDING 62 62 PYRIDOXAL PHOSPHATE.
SQ SEQUENCE 514 AA; 56276 MW; B36E5F26 CRC32;

Query Match 65.1%; Score 54; DB 1; Length 514;
Best Local Similarity 50.0%; Pred. No. 1.12e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 434 RLXSFEEFES 443
QY 62 KVVLFDFPEG 71

RESULT 13
ID THDI_ECOLI STANDARD; PRT; 514 AA.
AC P04968;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE THREONINE DEHYDRATASE BIOSYNTHETIC (EC 4.2.1.16) (THREONINE
DE DEAMINASE).
GN ILVA.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
CC [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 88036322.
RA COX J.L., COX B.J., FIDANZA V., CALHOUN D.H.;
RT "The complete nucleotide sequence of the ilvGMEDA cluster of
RT Escherichia coli K-12.";
RL GENE 56:185-198(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA GARRISON E., HARMS E., UMBARGER H.E.;
RL SUBMITTED (AUG-1986) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 87174741.
RA LAWTHER R.P., WEK R.C., LOPES J.M., PEREIRA R., TAILLON B.E.,
RA HATFIELD G.W.;
RT "The complete nucleotide sequence of the ilvGMEDA operon of
RT Escherichia coli K-12.";
RL NUCLEIC ACIDS RES. 15:2137-2155(1987).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 92358234.
RA DANIELS D.L., PLUNKETT G. III, BURLAND V.D., BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes.";
RL SCIENCE 257:771-778(1992).
RN [5]
RP SEQUENCE OF 439-514 FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 86111952.
RA WEK R.C., HATFIELD G.W.;
RT "Nucleotide sequence and in vivo expression of the ilvY and ilvC
RT genes in Escherichia coli K12. Transcription from divergent
RT overlapping promoters";
RL J. BIOL. CHEM. 261:2441-2450(1986).
RN [6]

SEQUENCE OF 1-10 FROM N.A.
RP RC STRAIN-K12;
RX MEDLINE; 89326124.
RA LOPES J.M., LAWTHER R.P.;
RT "Physical identification of an internal promoter, ilvAp, in the
RT distal portion of the ilvGMEDA operon.";
RL GENE 76:255-269(1989).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE; 98230745.
RA GALLAGHER D.T., GILLILAND G.L., XIAO G., ZONDLO J., FISHER K.E.,
RA CHINCHILLA D., EISENSTEIN E.;
RT "Structure and control of pyridoxal phosphate dependent allosteric
RT threonine deaminase.";
RL STRUCTURE 6:465-475(1998).
CC -1- FUNCTION: DEAMINATES L-THREONINE, BUT ALSO L-SERINE TO A LESSER
CC EXTENT.
CC -1- CATALYTIC ACTIVITY: L-THREONINE + H(2)O = 2-OXOBUTANOATE + NH(3)
CC + H(2)O.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHILE
CC VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME.
CC -1- PATHWAY: FIRST STEP OF ISOLEUCINE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: TO THREONINE DEHYDRATASE CATABOLIC AND SERINE
CC DEHYDRATASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; X04890; G288533; -.
EMBL; K03503; G146452; -.
EMBL; M10313; G146462; -.
EMBL; M11689; G146475; -.
EMBL; M32253; G146470; -.
EMBL; M87049; G148179; -.
EMBL; AE000453; G1790207; -.
EMBL; M25497; G538347; -.
PIR; B27310; DWEETS.
PIR; S30670; S30670.
PIR; S48895; S48895.
PIR; S48895; S48895.
PIR; LTDJ; 18-NOV-98.
ECO2DBASE; F050.1; 6TH EDITION.
RCOGENE; EG10493; ILVA.
PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
PFAM; PF00291; S_T_dehydratase; 1.
PFAM; PF00585; Thr_dehydrat_C; 2.
KW ISOLEUCINE BIOSYNTHESIS; LYASE; PYRIDOXAL PHOSPHATE;
KW ALLOSTERIC ENZYME; 3D-STRUCTURE.
FT BINDING 62 62 PYRIDOXAL PHOSPHATE.
FT CONFLICT 120 120 A -> R (IN REF. 2).
FT CONFLICT 140 140 A -> R (IN REF. 2).
FT CONFLICT 195 195 G -> C (IN REF. 2).
FT CONFLICT 243 243 A -> G (IN REF. 3).
FT CONFLICT 334 334 G -> V (IN REF. 2).
SQ SEQUENCE 514 AA; 56195 MW; 37AB88FA CRC32;

Query Match 65.1%; Score 54; DB 1; Length 514;
Best Local Similarity 50.0%; Pred. No. 1.12e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 434 RLXSFEEFES 443
QY 62 KVVLFDFPEG 71

RESULT 14
ID YVNF_AZOOCH STANDARD; PRT; 516 AA.
```

P24423;
AC 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE HYPOTHEICAL PROTEIN IN VNFED 5 REGION.
OS AZOTOBACTER CHROCOCCUM MCD 1.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; AZOTOBACTERACEAE;
OC AZOTOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90356423.
RA FALLIK E., ROBSON R.L.;
RT "Completed sequence of the region encoding the structural genes for
the vanadium nitrogenase of Azotobacter chroococcum.";
RL NUCLEIC ACIDS RES. 18:4616-4616(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51756; G38758; .
KW NITROGEN FIXATION; HYPOTHETICAL PROTEIN.
SQ SEQUENCE 516 AA; 58676 MW; B4B3A711 CRC32;

Query Match 65.1%; Score 54; DB 1; Length 516;
Best Local Similarity 83.3%; Pred. No. 1.12e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 457 VYLDEF 462
|||:|
Qy 63 VYLDEF 68

RESULT 15
ID THDH_ARXAD STANDARD; PRT; 550 AA.
AC 042615;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE THREONINE DEHYDRATASE PRECURSOR (EC 4.2.1.16) (THREONINE DEAMINASE).
GN ILV1.
OS ARXULA ADENINIVORANS.
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC CANDIDACEAE; ARXULA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=LS3;
RX MEDLINE; 98398049.
RA WARTMANN T., ROESSEL H., KUNZE I., BODE R., KUNZE G.;
RT "ILV1 gene from the yeast Arxula adeninivorans LS3 -- a new selective
transformation marker.";
RL YEAST 14:1017-1025(1998).
CC -!- CATALYTIC ACTIVITY: L-THREONINE + H(2)O = 2-OXOBUTANOATE + NH(3)
+ H(2)O.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- PATHWAY: FIRST STEP OF ISOLEUCINE BIOSYNTHESIS.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -!- SIMILARITY: TO THREONINE DEHYDRATASE CATABOLIC AND SERINE
DEHYDRATASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----

CC EMBL; AJ222772; E1215273; .
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW ISOLEUCINE BIOSYNTHESIS; LYASE; MITOCHONDRION; PYRIDOXAL PHOSPHATE;
FT TRANSIT PEPTIDE.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 550 THREONINE DEHYDRATASE.
FT BINDING 86 86 PYRIDOXAL PHOSPHATE.
SQ SEQUENCE 550 AA; 60416 MW; 1B06CB2F CRC32;

Query Match 65.1%; Score 54; DB 1; Length 550;
Best Local Similarity 55.6%; Pred. No. 1.12e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 468 RVFRFEPE 476
|||:|
Qy 62 KVLDFEPE 70

Search completed: Thu Jul 8 18:12:46 1999
Job time : 7 secs.

THIS PAGE BLANK (USPTO)

W P S R L A
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:13:06 1999; MasPar time 8.01 Seconds
Tabular output not generated. 68.115 Million cell updates/sec

Title: >US-09-041-236-2
Description: (62-71) from US09041236.pep (5 of 45)
Perfect Score: 83
Sequence: 1 KVLDFDFPEG 10

Scoring table: PAM i50
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprembl9
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.invertebrate 14:sp.virus

Statistics: Mean 25.426; Variance 40.341; scale 0.630

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	83	100.0	666	4	SEMAPHORIN L.	2.25e-04
2	68	81.9	393	11	SEMAPHORIN L. (FRAGMENT)	1.57e-01
3	66	79.5	775	2	POLYRIBONUCLEOTIDE NUC	3.58e-01
4	64	77.1	560	6	CYTOSOLIC IMP-GMP SPEC	8.08e-01
5	62	74.7	560	2	SERINE PROTEASE.	1.80e+00
6	60	72.3	160	14	GP19K PROTEIN.	3.95e+00
7	60	72.3	213	11	PHOSPHATIDYLINOSITOL S	3.95e+00
8	59	71.1	348	14	ENVELOPE GLYCOPROTEIN	5.83e+00
9	59	71.1	348	14	ENVELOPE GLYCOPROTEIN	5.83e+00
10	59	71.1	348	14	ENVELOPE GLYCOPROTEIN	5.83e+00
11	59	71.1	348	14	ENVELOPE GLYCOPROTEIN	5.83e+00
12	59	71.1	348	14	ENVELOPE GLYCOPROTEIN	5.83e+00
13	59	71.1	348	14	ENVELOPE GLYCOPROTEIN	5.83e+00
14	59	71.1	348	14	ENVELOPE GLYCOPROTEIN	5.83e+00
15	59	71.1	348	14	ENVELOPE GLYCOPROTEIN	5.83e+00
16	59	71.1	348	14	ENVELOPE GLYCOPROTEIN	5.83e+00
17	59	71.1	348	14	ENVELOPE GLYCOPROTEIN	5.83e+00
18	59	71.1	1124	5	F2392.4 PROTEIN.	5.83e+00
19	59	71.1	1244	14	POLYPROTEIN 2.	5.83e+00
20	59	71.1	1247	14	STRUCTURAL POLYPROTEIN	5.83e+00

21	58	69.9	88	14	P87625	41KBP FRAGMENT FROM LE	8.55e+00
22	58	69.9	93	14	O85338	HYPOTHETICAL 11.1 KD P	8.55e+00
23	58	69.9	194	5	O45270	C27H2.4.	8.55e+00
24	58	69.9	511	5	P90876	F46F3.2.	8.55e+00
25	58	69.9	1172	2	P94974	HYPOTHETICAL 128.2 KD	8.55e+00
26	58	69.9	1702	11	O54875	MYOTONIC DYSTROPHY KIN	8.55e+00
27	57	68.7	47	5	O23578	SIMILAR TO MARINER TRA	1.25e+01
28	57	68.7	93	14	O85351	HYPOTHETICAL 11.1 KD P	1.25e+01
29	57	68.7	380	5	O01448	SIMILAR TO STEROID/THY	1.25e+01
30	57	68.7	405	2	O31347	ORF1 PROTEIN.	1.25e+01
31	57	68.7	450	5	P91042	COSMID C10H11.	1.25e+01
32	56	67.5	73	2	O47834	ARPS.	1.82e+01
33	56	67.5	205	5	O17114	F39E9.5 PROTEIN.	1.82e+01
34	56	67.5	290	5	O61952	T24A6.16 PROTEIN.	1.82e+01
35	56	67.5	374	2	O66801	MANNOYLTRANSFERASE B.	1.82e+01
36	56	67.5	1201	10	O80649	T14N5.1 PROTEIN.	1.82e+01
37	55	66.3	135	14	O89505	ADENOVIRUS TYPE 40, CO	2.64e+01
38	55	66.3	221	14	O64824	REGULATORY PROTEIN ELA	2.64e+01
39	55	66.3	249	14	O64861	EIA (COMPLETE CDS) AND	2.64e+01
40	55	66.3	653	14	O64906	SIMILAR TO GENBANK ACC	2.64e+01
41	55	66.3	715	5	O23227	W09D12.1 PROTEIN.	2.64e+01
42	55	66.3	1207	2	P71068	YUKA PROTEIN.	2.64e+01
43	55	66.3	2523	2	O53393	PPE-FAMILY PROTEIN.	2.64e+01
44	55	66.3	2785	4	O75691	DRIM PROTEIN.	2.64e+01
45	54	65.1	1854	3	O14234	PROBABLE SODIUM CHANNE	3.81e+01

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	666 AA.
ID O75326	PRELIMINARY;	PRT;	666 AA.
AC O75326	PRELIMINARY;	PRT;	666 AA.
DT 01-NOV-1998	(TREMBLREL. 08, CREATED)		
DT 01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DT 01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE SEMAPHORIN L.			
GN SEMAL.			
OS HOMO SAPIENS (HUMAN).			
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC CATARRHINI; HOMINIDAE; HOMO.			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE; 98389619.			
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;			
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA			
RT viruses."			
RL GENOMICS 51:340-350(1998).			
DR EMBL; AF030698; G3523115; "			
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;			

Query Match 100.0%; Score 83; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 2.25e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	94	KVLFDFPEG 103
Qy	62	KVLFDFPEG 71

RESULT 2
ID O88371
AC O88371
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)

DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN L (FRAGMENT).

GN SEMAL.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389619.

RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses";
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030699; G3523117; -.
FT NON_TER 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;
Query Match 81.9%; Score 68; DB 11; Length 393;
Best Local Similarity 80.0%; Pred. No. 1.57e-01; Indels 0; Gaps 0;
Matches 8; Conservative 1; Mismatches 1;
Db 91 KYHFNPEPG 100
QY 62 KVLDFDPEPG 71
RESULT 3
ID O66593 PRELIMINARY; PRT; 775 AA.
AC O66593;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYRIBONUCLEOTIDE NUCLEOTIDYLTRANSFERASE.
GN PHPA.
OS AQUIFEX AEOLICUS.
OC BACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RC MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL NATURE 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE000679; G2982946; -.
KW TRANSFERASE.
SQ SEQUENCE 775 AA; 86488 MW; 52CA54B9 CRC32;
Query Match 79.5%; Score 66; DB 2; Length 775;
Best Local Similarity 80.0%; Pred. No. 3.58e-01; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 2;
Db 595 KYLFFVPPGG 604
QY 62 KVLDFDPEPG 71
RESULT 4
ID O46411 PRELIMINARY; PRT; 560 AA.
AC O46411;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CYTOSOLIC IMP-GMP SPECIFIC 5'-NUCLEOTIDASE.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=THYMUS;
RX MEDLINE; 98041889.
RA ALLEGRI S., PESI R., TOZZI M.G., FIOL C.J., JOHNSON R.B.,
RA ERIKSSON S.;

RT "Bovine cytosolic IMP/GMP-specific 5'-nucleotidase: cloning and
RT expression of active enzyme in Escherichia coli";
RL BIOCHEM. J. 328:483-487(1997).
DR EMBL; U73690; G2668557; -.
SQ SEQUENCE 560 AA; 64840 MW; DFA195B5 CRC32;
Query Match 77.1%; Score 64; DB 6; Length 560;
Best Local Similarity 70.0%; Pred. No. 8.08e-01; Indels 0; Gaps 0;
Matches 7; Conservative 1; Mismatches 2;
Db 260 MTYLDFDPHG 269
QY 62 KVLDFDPEPG 71
RESULT 5
ID O67764 PRELIMINARY; PRT; 560 AA.
AC O67764;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SERINE PROTEASE.
GN APRV.
OS AQUIFEX AEOLICUS.
OC BACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RC MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL NATURE 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE000764; G2984198; -.
KW PROTEASE.
SQ SEQUENCE 560 AA; 61463 MW; 81B6E25C CRC32;
Query Match 74.7%; Score 62; DB 2; Length 560;
Best Local Similarity 70.0%; Pred. No. 1.80e-00; Indels 0; Gaps 0;
Matches 7; Conservative 3; Mismatches 0;
Db 39 KYLLDLPDG 48
QY 62 KVLDFDPEPG 71
RESULT 6
ID O12399 PRELIMINARY; PRT; 160 AA.
AC O12399;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE GP19K PROTEIN.
GN AD1/E3-GP19K.
OS HUMAN ADENOVIRUS TYPE 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; ADENOVIRIDAE; MASTADENOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HANNOVER /ADRIAN;
RA REICHMANN H., SCHAARSCHMIDT E., GEISLER B., HAUSMANN J., ORTMANN D.,
RA BAUER U., FLUNKER G., SEIDEL W.,
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Y11260; E1191638; -.
DR EMBL; Y11032; E311262; -.

```
SQ SEQUENCE 160 AA; 18518 MW; 98903C66 CRC32;
Query Match 72.3%; Score 60; DB 14; Length 160;
Best Local Similarity 62.9%; Pred. No. 3.95e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 89 KVFLEYFP 96
QY 62 KVFLEDFP 69

RESULT 7
ID P70500 PRELIMINARY; PRT; 213 AA.
AC P70500;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE PHOSPHATIDYLINOSITOL SYNTHASE.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR;
RA HOSAKA K.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR;
RX MEDLINE; 96397537.
RA TANAKA S., NIKAWA J.I., IMAI H., YAMASHITA S.;
RT "Molecular cloning of rat phosphatidylinositol synthase cDNA by
RT functional complementation of the yeast Saccharomyces cerevisiae pis
RT mutation.";
RL FEBS LETT. 393:89-92(1996).
DR EMBL; D82928; D1012301; -.
DR PROSITE; PS00379; CDP-ALCOHOL_P_TRANSF; 1.
DR PFAM; PF01066; CDP-OH_P_TRANSF; 1.
SQ SEQUENCE 213 AA; 23613 MW; C4843649 CRC32;

Query Match 72.3%; Score 60; DB 11; Length 213;
Best Local Similarity 66.7%; Pred. No. 3.95e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 155 LYLFNFSEG 163
QY 63 VYLFDFPEG 71

RESULT 8
ID Q80742 PRELIMINARY; PRT; 348 AA.
AC Q80742;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN EI.
OS HIGHLANDS J VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
OC ALPHAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WX3-2AP;
RA CILNIS M.J., KANG W., WEAVER S.C.;
RL VIROLOGY 218:0-0(0).
DR EMBL; U52584; G1279808; -.
KW ENVELOPE PROTEIN.
FT NON_TER 1
SQ SEQUENCE 348 AA; 37381 MW; D7B2B08F CRC32;

Query Match 71.1%; Score 59; DB 14; Length 348;
Best Local Similarity 77.8%; Pred. No. 5.83e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 92 KVNVDPE 100
QY 62 KVFLEDFPE 70

RESULT 9
ID Q80746 PRELIMINARY; PRT; 348 AA.
AC Q80746;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN EI.
OS HIGHLANDS J VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
OC ALPHAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-F25;
RA CILNIS M.J., KANG W., WEAVER S.C.;
RL VIROLOGY 218:0-0(0).
DR EMBL; U52590; G1279820; -.
KW ENVELOPE PROTEIN.
FT NON_TER 1
SQ SEQUENCE 348 AA; 37353 MW; DF35ED81 CRC32;

Query Match 71.1%; Score 59; DB 14; Length 348;
Best Local Similarity 77.8%; Pred. No. 5.83e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 92 KVNVDPE 100
QY 62 KVFLEDFPE 70

RESULT 10
ID Q80745 PRELIMINARY; PRT; 348 AA.
AC Q80745;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN EI.
OS HIGHLANDS J VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
OC ALPHAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R362;
RA CILNIS M.J., KANG W., WEAVER S.C.;
RL VIROLOGY 218:0-0(0).
DR EMBL; U52589; G1279818; -.
KW ENVELOPE PROTEIN.
FT NON_TER 1
SQ SEQUENCE 348 AA; 37323 MW; A25CCF57 CRC32;

Query Match 71.1%; Score 59; DB 14; Length 348;
Best Local Similarity 77.8%; Pred. No. 5.83e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 92 KVNVDPE 100
QY 62 KVFLEDFPE 70

RESULT 11
ID Q80738 PRELIMINARY; PRT; 348 AA.
AC Q80738;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
```

```
GN OS HIGHLANDS J VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
OC ALPHAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WC-431;
RA CILNIS M.J., KANG W., WEAVER S.C.;
RL VIROLOGY 218:0-0(0).
DR EMBL; U52576; G1279792; -.
KW ENVELOPE PROTEIN.
FT NON_TER 1
SQ SEQUENCE 348 AA; 37353 MW; 19628FEC CRC32;

Query Match 71.1%; Score 59; DB 14; Length 348;
Best Local Similarity 77.8%; Pred. No. 5.83e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 92 KVINYDFPE 100
   ||| :||||
QY 62 KVLDFPE 70

RESULT 12
ID Q90205; PRELIMINARY; PRT; 348 AA.
AC Q90205;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN E1.
OS HIGHLANDS J VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
OC ALPHAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96193746.
RA CILNIS M.J., KANG W., WEAVER S.C.;
RL "Genetic Conservation of Highlands J viruses.";
RL VIROLOGY 218:343-351(1996).
DR EMBL; U52575; G1279790; -.
DR EMBL; U52577; G1279794; -.
DR EMBL; U52578; G1279796; -.
DR EMBL; U52579; G1279798; -.
DR EMBL; U52583; G1279806; -.
DR EMBL; U52586; G1279812; -.
DR EMBL; U52587; G1279814; -.
DR EMBL; U52591; G1279822; -.
DR EMBL; U52592; G1279824; -.
DR EMBL; U52593; G1279826; -.
KW ENVELOPE PROTEIN.
FT NON_TER 1
SQ SEQUENCE 348 AA; 37367 MW; ADE7891B CRC32;

Query Match 71.1%; Score 59; DB 14; Length 348;
Best Local Similarity 77.8%; Pred. No. 5.83e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 92 KVINYDFPE 100
   ||| :||||
QY 62 KVLDFPE 70

RESULT 13
ID Q80744; PRELIMINARY; PRT; 348 AA.
AC Q80744;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN E1.
OS HIGHLANDS J VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
```

```
OC ALPHAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3-81690;
RA CILNIS M.J., KANG W., WEAVER S.C.;
RL VIROLOGY 218:0-0(0).
DR EMBL; U52588; G1279816; -.
KW ENVELOPE PROTEIN.
FT NON_TER 1
SQ SEQUENCE 348 AA; 37297 MW; 1FBFFED3 CRC32;

Query Match 71.1%; Score 59; DB 14; Length 348;
Best Local Similarity 77.8%; Pred. No. 5.83e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 92 KVINYDFPE 100
   ||| :||||
QY 62 KVLDFPE 70

RESULT 14
ID Q80741; PRELIMINARY; PRT; 348 AA.
AC Q80741;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN E1.
OS HIGHLANDS J VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
OC ALPHAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCJ5-6X;
RA CILNIS M.J., KANG W., WEAVER S.C.;
RL VIROLOGY 218:0-0(0).
DR EMBL; U52582; G1279804; -.
KW ENVELOPE PROTEIN.
FT NON_TER 1
SQ SEQUENCE 348 AA; 37397 MW; 31F09C89 CRC32;

Query Match 71.1%; Score 59; DB 14; Length 348;
Best Local Similarity 77.8%; Pred. No. 5.83e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 92 KVINYDFPE 100
   ||| :||||
QY 62 KVLDFPE 70

RESULT 15
ID Q80739; PRELIMINARY; PRT; 348 AA.
AC Q80739;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN E1.
OS HIGHLANDS J VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
OC ALPHAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B-8-74;
RA CILNIS M.J., KANG W., WEAVER S.C.;
RL VIROLOGY 218:0-0(0).
DR EMBL; U52580; G1279800; -.
KW ENVELOPE PROTEIN.
FT NON_TER 1
SQ SEQUENCE 348 AA; 37395 MW; 1A59FCDA CRC32;

Query Match 71.1%; Score 59; DB 14; Length 348;
Best Local Similarity 77.8%; Pred. No. 5.83e+00;
```

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 92 KVNNDPPE 100

||| :|||

Qy 62 KVLDPPE 70

Search completed: Thu Jul 8 18:13:19 1999
Job time : 13 secs.

THIS PAGE BLANK (USPTO)

WQSEFH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:16:23 1999; MasPar time 11.43 Seconds
Tabular output not generated. 26.047 Million cell updates/sec

Title: >US-09-041-236-2
Description: (72-85) from US09041236.pep (6 of 45)
Perfect Score: 83
Sequence: 1 KNASVRTVNIGSTK 14

Scoring table:
PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 16.569; Variance 49.940; scale 0.332

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	51	61.4	738	9 R47899	Human melanoma-associ	6.53e+01
2	51	61.4	738	2 P70382	Sequence of the melan	6.53e+01
3	48	57.8	465	2 R11423	Tryptophanase.	1.37e+02
4	48	57.8	609	39 W86187	Vibriolysin, useful f	1.37e+02
5	48	57.8	609	4 R1414	Vibriolysin, a proteas	1.37e+02
6	48	57.8	610	1 P93578	Extracellular neutral	1.37e+02
7	48	57.8	792	10 R59765	Extracellular neutral	1.37e+02
8	46	55.4	838	34 W49078	Solanum tuberosum tub	2.23e+02
9	45	54.2	34	4 R20823	PND EE244-1.	2.84e+02
10	45	54.2	34	4 R20954	PND EE1476-1.	2.84e+02
11	45	54.2	34	4 R20825	PND EE244-3.	2.84e+02
12	45	54.2	34	4 R20824	PND EE244-2.	2.84e+02
13	45	54.2	372	39 W87637	Fusion protein His-ch	2.84e+02
14	45	54.2	357	11 R56479	Sorbitol dehydrogenas	2.84e+02
15	45	54.2	456	13 R74996	E. maxima Em70-1 anti	2.84e+02
16	45	54.2	943	34 W84378	Mycobacterium tubercu	2.84e+02

17	45	54.2	943	37 W81745	M. tuberculosis immun	2.84e+02
18	44	53.0	21	12 R60658	HTLV-SC SP-10 peptide	3.60e+02
19	44	53.0	21	37 W73239	SP-10 SC peptide.	3.60e+02
20	44	53.0	21	25 W16505	SP-10 SC region of HI	3.60e+02
21	44	53.0	21	16 R85362	HIV envelope glycopro	3.60e+02
22	44	53.0	29	16 R85375	HTLV-III SP-10-like p	3.60e+02
23	44	53.0	29	25 W16518	SP-10 SC region of HI	3.60e+02
24	44	53.0	29	37 W73248	SP-10 SC peptide.	3.60e+02
25	44	53.0	29	12 R83255	SP-10-like peptide (S	3.60e+02
26	44	53.0	31	32 W54920	HIV gp120 envelope pr	3.60e+02
27	44	53.0	32	14 R75830	HIV-1 isolate V3sc gp	3.60e+02
28	44	53.0	33	2 R11209	Retroviral B-epitope	3.60e+02
29	44	53.0	35	3 R14335	HIV-1 amplifiser pepi	3.60e+02
30	44	53.0	35	7 R36585	Virus neutralising ep	3.60e+02
31	44	53.0	39	25 W32959	C3-V3 SC immunogen co	3.60e+02
32	44	53.0	79	33 W61017	Streptococcus pneumon	3.60e+02
33	44	53.0	94	12 R67387	L. pneumophila GroES-	3.60e+02
34	44	53.0	231	1 R04496	HIV fusion protein PB	3.60e+02
35	44	53.0	231	1 P94806	PBisc HIV fusion prot	3.60e+02
36	44	53.0	344	20 W08097	Vmp7 soluble variant.	3.60e+02
37	44	53.0	374	29 W42435	Human promiscuous G-a	3.60e+02
38	44	53.0	855	32 W43069	HIV-1 gp120 protein f	3.60e+02
39	44	53.0	896	19 W00733	Human major vault pro	3.60e+02
40	44	53.0	25	2 R08275	HIV peptide fragment	4.56e+02
41	43	51.8	25	5 R31275	HIV principal determi	4.56e+02
42	43	51.8	35	4 R20894	PND EE505-2.	4.56e+02
43	43	51.8	599	39 W73479	Grapevine leafroll vi	4.56e+02
44	43	51.8	719	11 R60193	Modified protective a	4.56e+02
45	43	51.8	763	30 W41734	Human TRAF-2 kinase.	4.56e+02

ALIGNMENTS

RESULT 1
ID R47899 standard; Protein; 738 AA.
AC R47899;
DT 18-AUG-1994 (Revised)
DT 20-JUL-1994 (first entry)
DE Human melanoma-associated antigen p97.
KW Human melanoma-associated antigen p97; Melanotransferrin; GPI;
KW glycosyl-phosphatidylinositol anchor; truncated; soluble;
KW blood-brain barrier; iron binding; Alzheimer's disease;
KW iron metabolism; disorder; haemochromatosis; ischaemic tissue damage;
KW heart disease; skin cancer; brain tumour.
OS Synthetic.
FH Key Location/Qualifiers
FT peptide 1..19
FT protein 20..738
FT /label= p97
FT
FN W09401463-A.
PD 20-JAN-1994.
PF 09-JUL-1993; CA0272.
PR 10-JUL-1992; US-912291.
PR (UYBR-) UNIV BRITISH COLUMBIA.
PI Food MR, Jefferies WA, McGeer PL, Rothenberger S;
PI Yamada T;
PI WPI; 94-034993/04.
DR N-PSDB; Q55768.
DR New GPI-anchored and soluble forms of p97 - for treating
PT disorders of iron metabolism, delivering drugs to the brain and
PT for diagnosis and treatment of Alzheimer's disease
PS Disclosure; Page 102-106; 166pp; English.
CC The melanoma-associated antigen p97 has been found to be a
CC GPI-anchored protein expressed on the cell surface which is able to
CC bind iron. A soluble form of p97 can be isolated from the aqueous
CC phase after Triton-X-114 phase separation. The soluble (hydrophilic)
CC form of p97 does not contain ethanolamine and it has a slower rate
CC of transport than GPI-anchored p97. The role of p97 in iron-
CC its agonists, antagonists and stimulants may be useful in treatment
CC of conditions where iron-metabolism is disturbed, e.g.
CC haemochromatosis and skin cancer. Expression of p97 (and
CC transferrin receptors) by reactive microglial cells associated with

CC senile plaques in Alzheimer's Disease (AD) provides a means of
CC diagnosing AD.
SQ Sequence 738 AA;

Query Match 61.4%; Score 51; DB 9; Length 738;
Best Local Similarity 54.5%; Pred. No. 6.33e+01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 674 kdatvrvpvg 684

QY 72 KNASVRTVNI 82

RESULT 2

ID P70382 standard; protein; 738 AA.
AC P70382;
DT 19-FEB-1991 (first entry)
DE Sequence of the melanoma-associated P97 antigen.
KW Cancer vaccine: melanoma vaccine.
OS Homo sapiens.
PN NL8700285-A.
PD 01-SEP-1987.
PF 06-FEB-1987; 000285.
PR 07-FEB-1986; US-827313.
PS 27-JAN-1987; US-007230.
PA (ONCO-) ONCOGEN.
DR WPI: 87-275646/39.
DR N-PSDB: N70607.
PT Melanoma-associated antigen peptide(s) - are contained in
PT recombinant viruses for prodn. of vaccines against melanoma
PS Disclosure; Fig 3(A-B); 69pp; Dutch.
CC The sequence encoding the melanoma-associated P97 antigen and
CC recombinant viruses contg. it are useful for prodn. of vaccines
CC (claimed) against melanoma. Insertion of the P97 gene into vaccinia
CC virus yields the recombinant virus vp97a-WY. BSC cells infected with
CC this virus produce large ams. of P97, and the virus has
CC immunogenic activity in vivo.
SQ Sequence 738 AA;

Query Match 61.4%; Score 51; DB 2; Length 738;
Best Local Similarity 54.5%; Pred. No. 6.33e+01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 674 kdatvrvpvg 684

QY 72 KNASVRTVNI 82

RESULT 3

ID R11423 standard; Protein; 465 AA.
AC R11423;
DT 11-JUN-1991 (first entry)
DE Tryptophanase.
KW Tryptophanase.
OS Proteus inconstans IFO 12931.
PN J03047080-A.
PD 28-FEB-1991.
PF 13-JUL-1989; 181522.
PR (TOYM) TOYOB0 KK.
PA WPI: 91-105671/15.
DR N-PSDB: Q11248.
PT DNA sequence coding for tryptophanase - produced from Proteus
PT genus strain without adding tryptophan to culture
PS Disclosure; Fig 2; 7pp; Japanese.
CC The tryptophanase can be easily produced by recombinant techniques,
CC without the need to add tryptophan to the medium.
SQ Sequence 465 AA;

Query Match 57.8%; Score 48; DB 2; Length 465;
Best Local Similarity 54.5%; Pred. No. 1.37e+02;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 381 esgvraveigs 391
QY 73 NASVRTVNI 83

RESULT 4

ID W86187 standard; Protein; 609 AA.
AC W86187;
DT 29-MAR-1999 (first entry)
DE Vibriolysin, useful for wound healing and debridement.
KW Vibriolysin; neutral protease; debridement; wound healing; burn;
KW bone fracture; surgical abrasion; bed sore; ulcer; tendonitis;
KW bursitis; vaginitis; cervicitis; circumcision; episiotomy;
KW pilonidal cyst wart; carbuncle; sunburn; frostbite; therapy.
OS Vibrio proteolyticus strain ATCC 53559.
PN W09855604-A1.
PD 10-DEC-1998.
PF 01-JUN-1998; U10698.
PR 02-JUN-1997; US-867331.
PA (GRAC) GRACE & CO-CONN W R.
PI Durham DR, Fortney DZ, Yang K;
DR WPI: 99-070265/06.
DR N-PSDB: V84700.
PT Hydrophilic composition containing enzyme, especially protease, and
PT glyceryl cocoate - stable for months at room temperature,
PT specifically used to debride wounds and promote their healing
PS Disclosure; Page 20-25; 31pp; English.
CC This is the amino acid sequence of vibriolysin, a new extracellular
CC neutral protease of Vibrio proteolyticus ATCC 53559. Vibriolysin
CC can be obtained by fermentation of V. proteolyticus, or by
CC cultivation of recombinant host cells, collecting the enzyme from
CC the culture broth. A claimed hydrophilic pharmaceutical
CC composition comprises an extracellular neutral protease produced by
CC vibrio and enough glyceryl cocoate to maintain enzymatic activity
CC at over 80% for at least 100 days at room temperature. The
CC composition is used to debride wounds and to promote wound healing,
CC more generally to remove any necrotic and/or non-viable tissue.
CC Particular applications are treatment of burns, bone fractures,
CC surgical abrasions, bed sores, slowly healing ulcers, tendonitis,
CC bursitis, vaginitis, cervicitis, circumcision, episiotomy,
CC pilonidal cyst warts, carbuncles, sunburn and frostbite (claimed).
CC Vibriolysin is able to hydrolyse components of eschar (including
CC denatured collagen, elastin and fibrin) but does not damage native
CC tissue.
SQ Sequence 609 AA;

Query Match 57.8%; Score 48; DB 39; Length 609;
Best Local Similarity 57.1%; Pred. No. 1.37e+02;
Matches 8; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

Db 234 nnsavrtvdingst 247

QY 72 KNASVRTVNI-GST 84

RESULT 5

ID R21414 standard; Protein; 609 AA.
AC R21414;
DT 03-JUN-1992 (first entry)
DE Vibriolysin, a protease compsn. for wound treatment.
KW Protease; necrotic tissue; collagen; elastin; fibrin; wound
KW debridement; vascularisation.
OS Vibrio proteolyticus.
PN EP-472011-A.
PD 26-FEB-1992.
PF 29-JUL-1991; 112732.
PR 15-AUG-1990; US-567884.
PR 13-MAR-1991; US-670612.
PA (GRAC) GRACE W R & CO-CONN.
PI Fortney DZ, Durham DR;
DR WPI: 92-066263/09.
DR N-PSDB: Q21400.
PT Protease compsn. for wound treatment - contains protease from

PT Vibrio, esp. V proteolyticus, which hydrolyses components of
PT necrotic tissue but not native tissue
PS Disclosure: Fig 1: 23pp; English.
CC The vibriolysin protease can be produced by aerobic fermentation of
CC Vibrio species in nutrient medium. The protein sequence was
CC deduced from the DNA sequence obt'd. by cloning the Vibrio protease
CC from chromosomal DNA. Vibriolysin hydrolyses components of necrotic
CC tissue including denatured collagen, elastin and fibrin, while viable
CC living tissue is not injured. It causes wound debridement and
CC stimulates vascularisation and healing of traumatised tissue. The
CC protease can be used in management of full and partial thickness
CC wounds, burn wounds, debridement of ulcerative lesions, principally
CC pressure (decubitus) ulcers and varicose stasis and trophic ulcers,
CC prepn. of skin graft sites and general surgical wounds such as
CC amputation, incisional, traumatic and pyogenic wounds and for the
CC treatment of vaginitis, cervicitis, circumcisions, epistomy, cyst
CC wounds, carbuncles, sunburn, frostbite and cataract scar tissue.
SQ Sequence 609 AA;

Query Match 57.8%; Score 48; DB 4; Length 609;
Best Local Similarity 57.1%; Pred. No. 1.37e+02;
Matches 8; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

Db 234 nnsavrtvdingst 247
:|||||: |||
QY 72 KNASVRTVNI-GST 84

RESULT 6
ID P93578 standard; protein: 610 AA.
AC P93578;
DT 19-JUN-1990 (first entry)
DE Extracellular neutral protease (vibriolysin) as encoded by gene from
DE Vibrio proteolyticus ATCC 53559
KW Extracellular neutral protease; vibriolysin; Vibrio proteolyticus;
KW ATCC 53559.
OS Vibrio proteolyticus.
PN EP-309879-A.
PD 05-APR-1989.
PF 21-AUG-1988; 115439.
PR 19-NOV-1987; US-123038, US-103983.
PA (GRAC) Grace WR Co.
PI Deutch AH, David VA;
DR WPI: 89-101218/14.
DR N-PSDB; N90726 tag j.
PT Cloning and expression of neutral protease genes -
PT using DNA encoding enzymes of Vibrio proteolyticus or Bacillus
PT species in gram-negative microorganisms
PS Figure 1: 16pp; English.
CC Recombinant DNA contg. a Vibrio nuclear protease enzyme (NPE) gene is
CC claimed as are gram-negative microorganisms contg. DNA coding from more
CC than one NPE of Vibrio proteolyticus. The microorganisms are pref. E.coli
CC or Serratia species, esp. E.coli ATCC 67499 or 67501. The gram-negative
CC microorganisms synthesise active protease enzyme. The genes can be
CC manipulated for the overprdn. of NPE.
SQ Sequence 610 AA;

Query Match 57.8%; Score 48; DB 1; Length 610;
Best Local Similarity 57.1%; Pred. No. 1.37e+02;
Matches 8; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

Db 235 nnsavrtvdingst 248
:|||||: |||
QY 72 KNASVRTVNI-GST 84

RESULT 7
ID R59765 standard; protein: 792 AA.
AC R59765;
DT 20-FEB-1995 (first entry)
DE Extracellular neutral protease (vibriolysin).
KW Extracellular; neutral protease; vibriolysin; V. proteolyticus;
KW recombinant; gram-negative; microorganisms; E. coli; bacteria;

KW secretion; lysis.
OS Vibrio proteolyticus ATCC 53559.
FH Key Location/Qualifiers
FT misc_difference 1..793 /note= "x represents an in-frame stop codon"
FT protein 83..692 /note= "Encoded by large open reading frame"
FT
FT EP-605073-A.
PN 06-JUL-1994.
PD 21-SEP-1988; 115439.
PF 01-OCT-1987; US-103983.
PR 19-NOV-1987; US-123038.
PA (GRAC) GRACE & CO-CONN W R.
PI David VA, Deutch AH;
DR WPI: 94-210320/26.
DR N-PSDB; Q68538.
PT Recombinant DNA contg neutral protease genes - obt'd from Vibrio
PT species or Bacillus steartotherophilus, used for large scale
PT prodn of the enzyme.
PS Disclosure: Fig 1: 15pp; English.
CC This sequence represents the extracellular neutral protease (vibriolysin)
CC from V. proteolyticus ATCC 53559. The DNA encoding this sequence may be
CC used in the production of recombinant gram-negative microorganisms, such
CC as E. coli, which secrete the heterologous protease. The enzyme may also
CC be liberated by lysis of the microbial cell. Recombinant E. coli
CC producing the protease may be used for the large scale production of
CC active, functional, neutral protease enzymes for industrial use.
SQ Sequence 792 AA;

Query Match 57.8%; Score 48; DB 10; Length 792;
Best Local Similarity 57.1%; Pred. No. 1.37e+02;
Matches 8; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

Db 317 nnsavrtvdingst 330
:|||||: |||
QY 72 KNASVRTVNI-GST 84

RESULT 8
ID W49078 standard; protein: 838 AA.
AC W49078;
DT 09-NOV-1998 (first entry)
DE Solanum tuberosum tuber H-type alpha glucan phosphorylase.
KW H-type; tuber; alpha-glucan phosphorylase; cold storage; chips;
KW cold-sweetening.
OS Solanum tuberosum.
FH Key Location/Qualifiers
FT Peptide 1..838 /note= "mature peptide"
FT
FT W09835051-A1.
PN 13-AUG-1998.
PD 05-FEB-1998; CA0055.
PR 04-JUN-1997; US-868786.
PR 10-FEB-1997; US-036946.
PA (MIAC) CANADA DEPT AGRIC & AGRI-FOOD CANADA.
PI Armstrong JD, Kawchuk LM, Knowles NR, Lynch DR;
DR WPI: 98-447248/38.
DR N-PSDB; V32919.
PT Transgenic potato plants with improved cold storage characteristics
PT - are modified plants with reduced alpha glucan L or H type
PT phosphorylase activity, and have reduced cold sweetening
PS Disclosure: Page 58-61; 91pp; English.
CC The sequence is that of alpha-glucan H-type tuber phosphorylase
CC which can be used in the production of a potato plant with
CC improved tuber cold-storage characteristics. Such plants can be
CC used in agricultural potato production. The potatoes are especially
CC useful in the manufacture of potato chips. The method reduces
CC cold-sweetening (conversion of starches to sugars), resulting
CC in the ability to store potatoes cooler temperatures. This
CC produces prolonged dormancy, reduced incidence of disease and
CC increased storage life. The potatoes also have improved specific
CC gravity and improved chip score.
SQ Sequence 838 AA;

Query Match 55.4%; Score 46; DB 34; Length 838;
Best Local Similarity 58.3%; Pred. No. 2.23e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 600 knttptrvnmig 611
:|::|||::
QY 72 KNASVTVNIGSTK 85

RESULT 9

ID R20823 standard; Protein; 34 AA.
AC R20823;
DT 01-JUN-1992 (first entry)
DE PND EE244-1.
KW HIV; PND; principal neutralising determinant; Omp; Neisseria;
KW outer membrane proteosome; AIDS; vaccine; envelope gene.
OS Human immunodeficiency virus.
PN EP-471407-A.
PD 19-FEB-1992. 202025.
PF 07-AUG-1991; US-566656.
PR 13-AUG-1990; US-566638.
PR 13-AUG-1990; US-566654.
PA (MERI) MERCK & CO INC.
PI Lewis JA, Davide JP, Waterbury JA;
DR N-PSDB; Q21238.
PT New antigenic conjugate of HIV major neutralisation determinant -
PT covalently linked to outer membrane proteosome of Neisseria,
PT useful as vaccine and in treating and preventing HIV infections,
PT AIDS and ARC
PS Claim 1; Page 175 + 69; 177pp; English.
CC The sequences represented in Q21054-61, Q21194-311 and Q21321-78
CC encode PNDs of HIV (envelope protein) which are useful as immunogens
CC for AIDS vaccines, partic. in the form of conjugates.
CC To prepare antigenic conjugates, PND (of HIV) and Omp (of Neisseria)
CC are prep'd. separately, then linked by cross-linking agents, monogeneric
CC spacers or bigeneric spacers. The conjugate may comprise 1-50 PNDs.
CC A PND has 35 amino acids or less, but at least 5, it contains
CC Gly-X-Gly, where X is Pro, Leu, Ala, Gln or Ser.
CC The conjugates as AIDS or HIV vaccines can be used pre- or post-
CC exposure to prevent or treat HIV infection or disease, and are capable
CC of producing an immune response specific for the immunogen.
CC See also Q21052-61, Q21194-311 and Q21321-78.
SQ Sequence 34 AA;

Query Match 54.2%; Score 45; DB 4; Length 34;
Best Local Similarity 28.6%; Pred. No. 2.84e+02;
Matches 4; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 6 nnikirsihigpr 19
:|::|||::
QY 72 KNASVTVNIGSTK 85

RESULT 10

ID R20854 standard; Protein; 34 AA.
AC R20854;
DT 01-JUN-1992 (first entry)
DE PND EE1476-1.
KW HIV; PND; principal neutralising determinant; Omp; Neisseria;
KW outer membrane proteosome; AIDS; vaccine; envelope gene.
OS Human immunodeficiency virus.
PN EP-471407-A.
PD 19-FEB-1992. 202025.
PF 07-AUG-1991; US-566656.
PR 13-AUG-1990; US-566638.
PR 13-AUG-1990; US-566654.
PA (MERI) MERCK & CO INC.
PI Lewis JA, Davide JP, Waterbury JA;
DR WPI; 92-058471/08.

DR N-PSDB; Q21371.
PT New antigenic conjugate of HIV major neutralisation determinant -
PT covalently linked to outer membrane proteosome of Neisseria,
PT useful as vaccine and in treating and preventing HIV infections,
PT AIDS and ARC
PS Claim 1; Page 175 + 169; 177pp; English.
CC The sequences represented in Q21054-61, Q21194-311 and Q21321-78
CC encode PNDs of HIV (envelope protein) which are useful as immunogens
CC for AIDS vaccines, partic. in the form of conjugates.
CC To prepare antigenic conjugates, PND (of HIV) and Omp (of Neisseria)
CC are prep'd. separately, then linked by cross-linking agents, monogeneric
CC spacers or bigeneric spacers. The conjugate may comprise 1-50 PNDs.
CC A PND has 35 amino acids or less, but at least 5, it contains
CC Gly-X-Gly, where X is Pro, Leu, Ala, Gln or Ser.
CC The conjugates as AIDS or HIV vaccines can be used pre- or post-
CC exposure to prevent or treat HIV infection or disease, and are capable
CC of producing an immune response specific for the immunogen.
CC See also Q21052-61, Q21194-311 and Q21321-78.
SQ Sequence 34 AA;

Query Match 54.2%; Score 45; DB 4; Length 34;
Best Local Similarity 28.6%; Pred. No. 2.84e+02;
Matches 4; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 6 nnikirsihigpr 19
:|::|||::
QY 72 KNASVTVNIGSTK 85

RESULT 11

ID R20825 standard; Protein; 34 AA.
AC R20825;
DT 01-JUN-1992 (first entry)
DE PND EE244-3.
KW HIV; PND; principal neutralising determinant; Omp; Neisseria;
KW outer membrane proteosome; AIDS; vaccine; envelope gene.
OS Human immunodeficiency virus.
PN EP-471407-A.
PD 19-FEB-1992.
PF 07-AUG-1991; 202025.
PR 13-AUG-1990; US-566656.
PR 13-AUG-1990; US-566638.
PR 13-AUG-1990; US-566654.
PA (MERI) MERCK & CO INC.
PI Lewis JA, Davide JP, Waterbury JA;
DR WPI; 92-058471/08.
DR N-PSDB; Q21240.
PT New antigenic conjugate of HIV major neutralisation determinant -
PT covalently linked to outer membrane proteosome of Neisseria,
PT useful as vaccine and in treating and preventing HIV infections,
PT AIDS and ARC
PS Claim 1; Page 175 + 70-71; 177pp; English.
CC The sequences represented in Q21054-61, Q21194-311 and Q21321-78
CC encode PNDs of HIV (envelope protein) which are useful as immunogens
CC for AIDS vaccines, partic. in the form of conjugates.
CC To prepare antigenic conjugates, PND (of HIV) and Omp (of Neisseria)
CC are prep'd. separately, then linked by cross-linking agents, monogeneric
CC spacers or bigeneric spacers. The conjugate may comprise 1-50 PNDs.
CC A PND has 35 amino acids or less, but at least 5, it contains
CC Gly-X-Gly, where X is Pro, Leu, Ala, Gln or Ser.
CC The conjugates as AIDS or HIV vaccines can be used pre- or post-
CC exposure to prevent or treat HIV infection or disease, and are capable
CC of producing an immune response specific for the immunogen.
CC See also Q21052-61, Q21194-311 and Q21321-78.
SQ Sequence 34 AA;

Query Match 54.2%; Score 45; DB 4; Length 34;
Best Local Similarity 28.6%; Pred. No. 2.84e+02;
Matches 4; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 6 nnikirsihigpr 19
:|::|||::
QY 72 KNASVTVNIGSTK 85

```

RESULT 12
ID R20824 standard; Protein; 34 AA.
AC R20824;
DT 01-JUN-1992 (first entry)
DE PND E244-2.
KW HIV; PND; principal neutralising determinant; Omp; Neisseria;
OS outer membrane proteosome; AIDS; vaccine; envelope gene.
PN Human immunodeficiency virus.
EP-471407-A.
PD 19-FEB-1992.
PR 07-AUG-1991; 202025.
PR 13-AUG-1990; US-566656.
PR 13-AUG-1990; US-566638.
PR 13-AUG-1990; US-566654.
PA (MERI ) MERCK & CO INC.
PI Lewis JA, Davide JP, Waterbury JA;
DR WPI: 92-058471/08.
DR N-PSDB: Q21239.
PT New antigenic conjugate of HIV major neutralisation determinant -
PT covalently linked to outer membrane proteosome of Neisseria,
PT useful as vaccine and in treating and preventing HIV infections,
PT AIDS and ARC
PS Claim 1: Page 175 + 70; 177pp; English.
CC The sequences represented in Q21054-61, Q21194-311 and Q21321-78
CC encode PNDs of HIV (envelope protein) which are useful as immunogens
CC for AIDS vaccines, partic. in the form of conjugates.
CC To prepare antigenic conjugates, PND (of HIV) and Omp (of Neisseria)
CC are prepd. separately, then linked by cross-linking agents, monogeneric
CC spacers or bigeneric spacers. The conjugate may comprise 1-50 PNDs.
CC A PND has 35 amino acids or less, but at least 5, it contains
CC Gly-X-Gly, where X is Pro, Leu, Ala, Gln or Ser.
CC The conjugates as AIDS or HIV vaccines can be used pre- or post-
CC exposure to prevent or treat HIV infection or disease, and are capable
CC of producing an immune response specific for the immunogen.
CC See also Q21052-61, Q21194-311 and Q21321-78.
SQ Sequence 34 AA;

Query Match 54.2%; Score 45; DB 4; Length 34;
Best Local Similarity 28.6%; Pred. No. 2.84e+02;
Matches 4; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 6 nnikirsihigpar 19
QY 72 KNASVRTVNIIGSTK 85

RESULT 13
ID W87637 standard; Protein; 272 AA.
AC W87637;
DT 09-MAR-1999 (first entry)
DE Fusion protein His-chymosin pro-peptide-carp growth hormone.
KW Fusion protein; bovine chymosin pro-peptide; leech hirudin;
KW anticoagulant protein; autocatalytically maturing zymogen;
KW carp growth hormone; vaccine.
OS Synthetic.
OS Bos sp.
OS Cyprinus carpio.
FH Key Location/Qualifiers
FT Region 5..10
FT /note= "poly Histidine site"
FT Binding_site 27..31
FT /note= "enterokinase recognition site"
FT Cleavage_site 31..32
FT /note= "enterokinase cleavage site"
FT Peptide 43..84
FT /note= "bovine chymosin pro-peptide"
FT Protein 85..261
FT /note= "carp growth hormone"
FN W09849326-A1.
PD 05-NOV-1998.
PF 23-APR-1998; CA0398.

```

```

PR 25-APR-1997; US-044254.
PA (SEMB-) SEMBIOSYS GENETICS INC.
PI Alcantara J, Moloney M, Van Rooijen G;
DR WPI: 99-059646/05.
DR N-PSDB; V83967.
PT Preparation of recombinant polypeptides - by expression of a fusion
PT protein comprising a pro-peptide derived from an autocatalytically
PT maturing zymogen and a heterologous polypeptide
PS Example 2; Fig 2; 44pp; English.
CC The present sequence represents a fusion protein comprising a His
CC tag-bovine chymosin pro-peptide-carp growth hormone. The fusion
CC protein was made to exemplify the invention. The specification describes
CC a method for preparing a recombinant polypeptide in a host cell. A
CC chimeric nucleic acid sequence encoding a fusion protein comprising
CC a pro-peptide derived from an autocatalytically maturing zymogen linked
CC to a protein heterologous to the pro-peptide, is introduced into the host
CC cell. The host cells are then grown to produce the fusion protein.
CC Altering the environment of the fusion protein allows cleavage of the
CC pro-peptide from the fusion protein to release the recombinant
CC polypeptide. The method can be used for the preparation of recombinant
CC polypeptides such as hirudin or carp growth hormone. The fusion
CC proteins can be used for delivering to a human or animal a therapeutic
CC or nutritional polypeptide such as a vaccine, a peptide antibiotic, a
CC cattle feed enzyme, a cytokine, a gastric lipase or a lactase.
SQ Sequence 272 AA;

Query Match 54.2%; Score 45; DB 39; Length 272;
Best Local Similarity 46.2%; Pred. No. 2.84e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 31 kdpssrsaeigsa 43
QY 72 KNASVRTVNIIGST 84

RESULT 14
ID R56479 standard; protein; 357 AA.
AC R56479;
DT 07-MAR-1995 (first entry)
DE Sorbitol dehydrogenase from Saccharomyces cerevisiae.
KW Sorbitol dehydrogenase; Yeast regulatory sequence; assay reagent;
KW heterologous protein expression; superoxide dismutase; antibody;
KW hepatitis delta antigen.
OS Saccharomyces cerevisiae.
PN W09415942-A.
PD 21-JUL-1994.
PF 22-DEC-1993; U12560.
PR 30-DEC-1992; US-998226.
PA (ABBO ) ABBOTT LAB.
PI Sarthy AV, Schopp CW;
DR WPI: 94-249120/30.
DR N-PSDB; Q66579.
PT Purified yeast sorbitol dehydrogenase - used for isolating gene
PT regulatory sequences for use in expressing heterologous
PT polypeptide(s) in yeast
PS Claim 27; Page 31-34; 60pp; English.
CC This sequence (R56479) is that of sorbitol dehydrogenase.
CC The presence of eleven amino acids of the sorbitol dehydrogenase
CC from its N-terminal in a fusion polypeptide results in several fold
CC higher levels of expression of the fused polypeptides as compared to
CC the unfused polypeptide. The DNA encoding the enzyme is shown in
CC Q66579, it is useful for its regulatory elements. These can be used
CC in yeast hosts for the production of heterologous proteins, eg.
CC superoxide dismutase or the hepatitis delta antigen.
SQ Sequence 357 AA;

Query Match 54.2%; Score 45; DB 11; Length 357;
Best Local Similarity 38.5%; Pred. No. 2.84e+02;
Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 260 daavkttkvgtm 272
QY 73 NASVRTVNIIGSTK 85

```

```

RESULT 15
ID R74996 standard; Protein: 456 AA.
AC R74996;
DT 07-NOV-1995 (first entry)
DE E. maxima Em70-1 antigen.
KW T-lymphocyte stimulatory protein; coccidiosis; vector; vaccine;
KW poultry; immunogen; p70; Em70-1; antigen..
OS Eimeria maxima.
PN EP-653489-A.
PD 17-MAY-1995.
PF 09-NOV-1994; 308257.
PR 12-NOV-1993; EP-309078.
PA (ALKU ) AKZO NOBEL NV.
PI Bumstead JM, Dunn PPJ, Tomley FM, Vermeulen AN;
DR WPI: 95-180711/24.
DR N-PSDB; Q90495.
PT New Eimeria T-lymphocyte stimulatory protein - used to develop prods.
PT for the diagnosis, prevention and therapy of coccidiosis in poultry.
PS Disclosure; Page 21-23; 37pp; English.
CC A 70 kDa (p70) protein obtained from sporozoites isolated from E.
CC maxima-infected chickens was used to raise antibodies in rabbits.
CC The antiserum was used to screen a cDNA library of E. maxima
CC sporozoites in lambda ZAPII. Clone pEm70-1 had the sequence given
CC in Q90495. The Em70-1 antigen (R74996) was expressed in E. coli
CC and Sf9 insect cells.
SQ Sequence 456 AA;

Query Match 54.2%; Score 45; DB 13; Length 456;
Best Local Similarity 36.4%; Pred. No. 2.84e+02;
Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 131 kdaniridfg 141
OY 72 KNASVRTVNIG 82

Search completed: Thu Jul 8 18:16:42 1999
Job time : 19 secs.

```



```

6
RESULT      6
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM    #type complete
DATE        phosphotransferase system enzyme I (EC 2.7.3.9) - Bacillus
            subtilis
            #formal_name Bacillus subtilis
            #cross-references GB:M98359; NID:g1943574; PID:g1943575
            #note      sequence extracted from NCBI backbone (NCBIN:134327,
            NCBI:P134330)
ACCESSIONS  A46238; A47084; S04178; G69683
REFERENCE   A46238
#authors    M.H.
#journal    Protein Sci. (1993) 2:506-521
#title      Sequence analyses and evolutionary relationships among the
            energy-coupling proteins Enzyme I and HPr of the bacterial
            phosphoenolpyruvate: sugar phosphotransferase system.
            #cross-references MUID:93299364
            #accession C46238
            #molecule_type DNA
            #residues 1-67,88-570 #label REI1
            #cross-references GB:M98359; NID:g1943574; PID:g1943575
            #note      sequence extracted from NCBI backbone (NCBIN:134327,
            NCBI:P134330)
REFERENCE   A47084
#authors    Fajardo-Cavazos, P.; Salazar, C.; Nicholson, W.L.
#journal    J. Bacteriol. (1993) 175:1735-1744
#title      Molecular cloning and characterization of the Bacillus
            subtilis spore photoproduet lyase (spl) gene, which is
            involved in repair of UV radiation-induced DNA damage
            during spore germination.
            #cross-references MUID:93194799
            #accession A47084
            #molecule_type DNA
            #residues 398-570 #label FAJ
            #cross-references GB:L08809; NID:g289267; PID:g289268
            #experimental_source strain 168
            #note      sequence extracted from NCBI backbone (NCBIN:127497,
            NCBI:P127499)
REFERENCE   S04178
#authors    Gonzy-Treboul, G.; Zagorec, M.; Rain-Guion, M.C.; Steinmetz,
            M.
#journal    Mol. Microbiol. (1989) 3:103-112
#title      Phosphoenolpyruvate:sugar phosphotransferase system of
            Bacillus subtilis: nucleotide sequence of ptsA, ptsH and
            the 5'-end of ptsI and evidence for a ptsHI operon.
            #cross-references MUID:89237891
            #accession S04178
            #molecule_type DNA
            #residues 1-102 #label GON
            #cross-references EMBL:X12832; NID:g48679; PID:g48683
            A69580
REFERENCE   A69580
#authors    Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
            Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
            Boloitin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
            A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
            Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
            Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
            Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
            Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
            Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
            M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
            S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
            Guiseppe, G.; Guy, B.J.; Haga, K.; Hatech, J.; Harwood,
            C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
            Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
            Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
            Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
            Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
            Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
            Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
            M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
            M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
            V.; Pohl, T.M.; Portetelie, D.; Porwolnik, S.; Prescott,

```

```

A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viarri, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzensgager, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
#journal    Nature (1997) 390:249-256
#title      The complete genome sequence of the Gram-positive bacterium
            Bacillus subtilis.
            #cross-references MUID:98044033
            #accession G69683
            #status      nucleic acid sequence not shown; translation not shown
            #molecule_type DNA
            #residues 1-67,88-570 #label KUN
            #cross-references GB:Z99111; GB:AL009126; NID:g2633699; PID:el184981;
            PID:g2633762
            #experimental_source strain 168
GENETICS
#gene      ptsI
CLASSIFICATION #superfamily phosphotransferase system enzyme I;
               phosphotransferase system enzyme I homology
KEYWORDS      phosphoprotein; phosphotransferase; sugar transport system
FEATURE
4-561         #domain phosphotransferase system enzyme I homology
               #label PII
SUMMARY       #length 570 #molecular-weight 63079 #checksum 8196
               Query Match 62.7%; Score 52; DB 1; Length 570;
               Best Local Similarity 50.0%; Pred. No. 2.95e+00;
               Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Db 327 KSVVVRTLDIGGDK 340
I: |||:|:|:|
QY 72 KNASVRTVNIIGSTK 85
RESULT 7
ENTRY   WZEC #type complete
TITLE   tryptophanase (EC 4.1.99.1) - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE      14-Nov-1983 #sequence_revision 05-Dec-1997 #text_change
ACCESSION E65173; A91789; I41097; I69358; A01136
REFERENCE A64720
#authors  Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
            Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
            Rude, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
            Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
            Y.
#journal    Science (1997) 277:1453-1462
#title      The complete genome sequence of Escherichia coli K-12.
            #cross-references MUID:97456617
            #accession E65173
            #status      nucleic acid sequence not shown; translation not shown
            #molecule_type DNA
            #residues 1-476 #label BLAT
            #cross-references GB:AE000448; GB:U00096; NID:g1790142; PID:g1790144;
            UWGP:b3708
            #experimental_source strain K-12, substrain MG1655
REFERENCE   A91789
#authors    Deeley, M.C.; Yanofsky, C.
#journal    J. Bacteriol. (1981) 147:787-796
#title      Nucleotide sequence of the structural gene for tryptophanase
            of Escherichia coli K-12.
            #cross-references MUID:82007678
            #accession A91789

```

```
##molecule_type DNA
##residues 6-141,'T',143,'QG',146-383,'TG',386-476 ##label DEE
##experimental_source strain K-12
##note the authors translated the codons GAT, ACG, and CAG for
##note residues 142, 144, and 145 as Thr, Gln, and Gly,
##note respectively
REFERENCE
#authors A92100
#journal Kagamiyama, H.; Matsubara, H.; Snell, E.E.
#journal J. Biol. Chem. (1972) 247:1576-1586
#title The chemical structure of tryptophanase from Escherichia
#title coli. III. Isolation and amino acid sequence of the tryptic
#title peptides.
#cross-references MUID:72134434
#contents annotation: sequences of tryptic peptides; strain K-12
#note Lys-275 binds pyridoxal 5'-phosphate
REFERENCE
#authors I41097
#authors Tokushige, M.; Tsujimoto, N.; Oda, T.; Honda, T.; Yumoto, N.;
#journal Ito, S.; Yanamoto, M.; Kim, E.H.; Hiragi, Y.
#journal Biochimie (1989) 71:711-720
#title Role of cysteine residues in tryptophanase for monovalent
#title cation-induced activation.
#cross-references MUID:89323226
#accession I41097
#status translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 6-383,'TG',386-476 ##label RES
##residues 6-383,'TG',386-476 ##label RES
#cross-references EMBL:X15974; NID:g41935; PID:g41936
REFERENCE
#authors I54862
#authors Stewart, V.J.; Yanofsky, C.
#journal J. Bacteriol. (1985) 164:731-740
#title Evidence for transcription antitermination control of
#title tryptophanase operon expression in Escherichia coli K-12.
#cross-references MUID:86033634
#accession I69358
#status translated from GB/EMBL/DBJ
##molecule_type DNA
#residues 6-27 ##label RE2
#cross-references GB:M1190; NID:g147999; PID:g148001
GENETICS
#gene tnaA
#map_position 83 min
#complex homotetramer
#description catalyzes the degradation of tryptophan to indole, pyruvate,
#description and ammonia; it also catalyzes the synthesis of tryptophan
#description from indole and serine
CLASSIFICATION
#superfamily tryptophanase
KEYWORDS
#carbon-carbon lyase; homotetramer
SUMMARY
#length 476 #molecular_weight 53410 #checksum 8080
Query Match 61.4%; Score 51; DB 1; Length 476;
Best Local Similarity 58.3%; Pred. No. 4.78e+00;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 392 KVAGIRAVEIGS 403
| |::|:|:|
Qy 72 KNASVTVNIGS 83
#gene
#accession GDB:MFI2
#cross-references GDB:119387; OMIM:155750
#map_position 3q28-3q29
#classification superfamily transferrin; transferrin repeat homology
#keywords duplication; glycoprotein; lipoprotein; membrane protein;
#keywords metal binding; phosphatidylinositol linkage; phosphoprotein
FEATURE
#domain signal sequence #status predicted #label SIG\
#domain transferrin repeat homology #label TRH1\
#product melanotransferrin #status predicted #label MTF\
#domain transferrin repeat homology #label TRH2\
#domain carboxyl-terminal propeptide #status predicted
#label CTP\
38,135,515 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
```

```
#cross-references MUID:91358378
#accession B38120
#status preliminary
##molecule_type DNA
##residues 1-592 ##label PRI
#cross-references GB:M69036; NID:g141961; PID:g141963
CLASSIFICATION #superfamily phosphotransferase system enzyme I;
#keywords phosphotransferase system enzyme I homology
#keywords phosphoprotein
FEATURE
#domain phosphotransferase system enzyme I homology
#label PRI
SUMMARY
#length 592 #molecular_weight 65207 #checksum 8444
Query Match 61.4%; Score 51; DB 2; Length 592;
Best Local Similarity 45.5%; Pred. No. 4.78e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 343 TIRTDIGADK 353
|:|:|:|:|
Qy 75 SVRTVNIGSTK 85
#status translated from GB/EMBL/DBJ
##molecule_type mRNA
#residues 1-738 ##label ROS
#cross-references EMBL:M12154; NID:g189515; PID:g189518
#experimental_source melanoma
REFERENCE
#accession A60925
#authors Furukawa, K.S.; Furukawa, K.; Real, F.X.; Old, L.J.; Lloyd,
#journal K.O.
#journal J. Exp. Med. (1989) 169:585-590
#title A unique antigenic epitope of human melanoma is carried on
#title the common melanoma glycoprotein gp95/p97.
#cross-references MUID:89094252
#accession A60925
##molecule_type protein
#residues 20-25,'X',27-28,'X',30 ##label FUR
#comment This protein is found predominantly in human melanomas and in
#comment certain fetal tissues; it is found in only trace amounts in
#comment normal adult tissues.
#comment Seven disulfide bonds are predicted in each domain.
GENETICS
#gene GDB:MFI2
#cross-references GDB:119387; OMIM:155750
#map_position 3q28-3q29
#classification superfamily transferrin; transferrin repeat homology
#keywords duplication; glycoprotein; lipoprotein; membrane protein;
#keywords metal binding; phosphatidylinositol linkage; phosphoprotein
FEATURE
#domain signal sequence #status predicted #label SIG\
#domain transferrin repeat homology #label TRH1\
#product melanotransferrin #status predicted #label MTF\
#domain transferrin repeat homology #label TRH2\
#domain carboxyl-terminal propeptide #status predicted
#label CTP\
38,135,515 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
```



```
710      #modified site GPI-anchor ethanolamine amidated carboxyl
      end (Ser) (in mature form) #status predicted
SUMMARY  #length 738  #molecular-weight 80241  #checksum 9466

Query Match      61.4%  Score 51;  DB 1;  Length 738;
Best Local Similarity 54.5%;  Pred. No. 4.78e+00;
Matches 6;  Conservative 4;  Mismatches 1;  Indels 0;  Gaps 0;

Db  674 KDATVPVPG 584
   |::|::|::|
QY  72 KNASVRTVNI 82

RESULT 10
ENTRY   B42374      #type complete
TITLE   phosphotransferase system enzyme I (EC 2.7.3.9) -
        Staphylococcus carnosus
ORGANISM #formal_name Staphylococcus carnosus
DATE     10-Jul-1992 #sequence_revision 31-Jan-1997 #text_change
        23-Feb-1997
ACCESSION B42374; S17075
REFERENCE #authors Kohlbrecher, D.; Eisermann, R.; Hengstenberg, W.
          #journal J. Bacteriol. (1992) 174:2208-2214
          #title Staphylococcal phosphoenolpyruvate-dependent
        phosphotransferase system: molecular cloning and nucleotide
        sequence of the Staphylococcus carnosus ptsI gene and
        expression and complementation studies of the gene product.
        #cross-references MUID:92202148
        #accession B42374
        ##status preliminary
        ##molecule_type DNA
        ##residues 1-574 #label KOH
        ##cross-references GB:M69050; NID:g153073; PID:g153075
REFERENCE S15367
#authors Eisermann, R.; Fischer, R.; Kessler, U.; Neubauer, A.;
        Hengstenberg, W.
#journal Eur. J. Biochem. (1991) 197:9-14
#title Staphylococcal phosphoenolpyruvate-dependent
        phosphotransferase system. Purification and protein
        sequencing of the Staphylococcus carnosus
        histidine-containing protein, and cloning and DNA
        sequencing of the ptsH gene.
        #cross-references MUID:91200066
        #accession S17075
        ##status preliminary
        ##molecule_type DNA
        ##residues 1-90 #label EIS
        ##cross-references EMBL:X60766; NID:g46907; PID:g46909
CLASSIFICATION #superfamily phosphotransferase system enzyme I;
        phosphotransferase system enzyme I homology
        phosphoprotein; phosphotransferase; sugar transport system
KEYWORDS 5-563
FEATURE   #domain phosphotransferase system enzyme I homology
          #label Ptl
SUMMARY  #length 574  #molecular-weight 63369  #checksum 7968

Query Match      60.2%  Score 50;  DB 1;  Length 574;
Best Local Similarity 50.0%;  Pred. No. 7.68e+00;
Matches 7;  Conservative 3;  Mismatches 4;  Indels 0;  Gaps 0;

Db  329 KRVVRTLIGGDK 342
   |::|::|::|
QY  72 KNASVRTVNI 85

RESULT 11
ENTRY   Jc1375      #type complete
TITLE   phosphotransferase system enzyme I (EC 2.7.3.9) -
        Streptococcus salivarius
ORGANISM #formal_name Streptococcus salivarius
ALTERNATE_NAMES phosphoenolpyruvate--sugar phosphotransferase transport
        system enzyme I
        #formal_name Streptococcus salivarius
```

```
DATE     30-Sep-1993 #sequence_revision 31-Jan-1997 #text_change
        23-Feb-1997
ACCESSION Jc1375
REFERENCE #authors Gagnon, G.; Vadeboncoeur, C.; Levesque, R.C.; Frenette, M.
          #journal Gene (1992) 121:71-78
          #title Cloning, sequencing and expression in Escherichia coli of the
        ptsI gene encoding enzyme I of the
        phosphoenolpyruvate:sugar phosphotransferase transport
        system from Streptococcus salivarius.
        #cross-references MUID:93051364
        #accession Jc1375
        ##molecule_type DNA
        ##residues 1-577 #label GAG
        ##cross-references GB:M81756; NID:g153614; PID:g153615
GENETICS
#gene     ptsI
CLASSIFICATION #superfamily phosphotransferase system enzyme I;
        phosphotransferase system enzyme I homology
        phosphoprotein; phosphotransferase; sugar transport system
KEYWORDS 5-565
FEATURE   #domain phosphotransferase system enzyme I homology
          #label Ptl
SUMMARY  #length 577  #molecular-weight 63046  #checksum 1035

Query Match      60.2%  Score 50;  DB 1;  Length 577;
Best Local Similarity 50.0%;  Pred. No. 7.68e+00;
Matches 7;  Conservative 3;  Mismatches 4;  Indels 0;  Gaps 0;

Db  329 KPVVVRTMDIGDK 342
   |::|::|::|
QY  72 KNASVRTVNI 85

RESULT 12
ENTRY   S53639      #type complete
TITLE   protein kinase clk3 (EC 2.7.1.1) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change
        10-Jul-1998
ACCESSION S53639; S71040
REFERENCE #authors Hanes, J.; von der Kammer, H.; Klaudiny, J.; Scheit, K.H.
          #journal J. Mol. Biol. (1994) 244:665-672
          #title Characterization by cDNA cloning of two new human protein
        kinases. Evidence by sequence comparison of a new family of
        mammalian protein kinases.
        #cross-references MUID:95082033
        #accession S53639
        ##molecule_type mRNA
        ##residues 1-490 #label HAN
        ##cross-references GB:L29217
REFERENCE S71040
#authors Hanes, J.J.; der Kammer, H.; Klaudiny, J.J.; Scheit, K.H.
#submission submitted to the EMBL Data Library, January 1995
#accession S71040
        ##molecule_type mRNA
        ##residues 1-131,134-490 #label HAW
        ##cross-references EMBL:L29217; NID:g632971; PID:g632972
CLASSIFICATION #superfamily protein kinase homology
        alternative splicing; phosphotransferase
KEYWORDS 154-433
FEATURE   #domain protein kinase homology #label KIN
          #length 490  #molecular-weight 58588  #checksum 391
SUMMARY

Query Match      59.0%  Score 49;  DB 2;  Length 490;
Best Local Similarity 46.2%;  Pred. No. 1.23e+01;
Matches 6;  Conservative 5;  Mismatches 2;  Indels 0;  Gaps 0;

Db  312 KNTSIRVADFGSA 324
   |::|::|::|
QY  72 KNASVRTVNI 84
```

```

RESULT 13
ENTRY
TITLE      S70352      #type complete
ALTERNATE_NAMES  protein kinase clk3 (EC 2.7.1.-), testis-specific - rat
ORGANISM    LAMMER kinase
DATE        #formal_name Rattus norvegicus #common_name Norway rat
            21-Apr-1997 #sequence_revision 09-May-1997 #text_change
            10-Jul-1998
ACCESSIONS  S70352
REFERENCE    Becker, W.; Kentrup, H.; Heukelbach, J.; Joost, H.G.
#authors
#journal    Biochim. Biophys. Acta (1996) 1312:63-67
#title      cDNA cloning and characterization of rat Clk3, a LAMMER
            kinase predominantly expressed in testis.
#cross-references MUID:96271481
#accession    S70352
#molecule_type mRNA
#residues     1-490 #label BEC
#cross-references EMBL:X94351; NID:g1149536; PID:e217388; PID:g1149537
GENETICS
#gene         clk3
CLASSIFICATION #superfamily protein kinase homology
KEYWORDS      phosphotransferase; protein kinase
FEATURE       154-433
SUMMARY       #domain protein kinase homology #label KIN
            #length 490 #molecular-weight 58501 #checksum 9679
Query Match      59.0%; Score 49; DB 2; Length 490;
Best Local Similarity 46.2%; Pred. No. 1.23e+01;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
Db 312 KNTSIRVADFGSA 324
||:|:| :||:
Qy 72 KNASVRTVNI GST 84

RESULT 14
ENTRY
TITLE      A54099      #type complete
ALTERNATE_NAMES  protein kinase Darkener-of-apricot (EC 2.7.1.-) - fruit fly
            (Drosophila melanogaster)
ORGANISM    LAMMER protein kinase Doa
DATE        #formal_name Drosophila melanogaster
            06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
            24-Sep-1998
ACCESSIONS  A54099; S44077
REFERENCE    Yun, B.; Farkas, R.; Lee, K.; Rabinow, L.
#authors     Genes Dev. (1994) 8:1160-1173
#journal     The Doa locus encodes a member of a new protein kinase family
            and is essential for eye and embryonic development in
            Drosophila melanogaster.
#cross-references MUID:95011531
#accession    A54099
#status       preliminary
#molecule_type mRNA
#residues     1-517 #label YUN
#cross-references GB:X78715; NID:g472912; PID:g472913
GENETICS
#gene         FlyBase:Doa
CLASSIFICATION #cross-references FlyBase:FBgn0000480
KEYWORDS      #superfamily protein kinase homology
            ATP; autophosphorylation; phosphotransferase; protein kinase
FEATURE       168-438
SUMMARY       #domain protein kinase homology #label KIN
            #length 517 #molecular-weight 60975 #checksum 5562
Query Match      59.0%; Score 49; DB 2; Length 517;
Best Local Similarity 46.2%; Pred. No. 1.23e+01;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
Db 325 KNTDVRLLDFGSA 337
||:|:| :||:
Qy 72 KNASVRTVNI GST 84

RESULT 15
ENTRY
TITLE      F70556      #type complete
ALTERNATE_NAMES  hypothetical protein Rv1165 - Mycobacterium tuberculosis
            (strain H37RV)
ORGANISM    #formal_name Mycobacterium tuberculosis
            17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
            16-Dec-1998
ACCESSIONS  F70556
REFERENCE    Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
            C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry
            III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
            Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
            Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
            Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
            Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
            Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
            Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
            Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal     Nature (1998) 393:537-544
#title      Deciphering the biology of Mycobacterium tuberculosis from
            the complete genome sequence.
#cross-references MUID:98295987
#accession    F70556
#status       preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues     1-628 #label COL
#cross-references GB:Z95584; GB:AL123456; NID:g3261774; PID:e1299888;
            PID:g3261775
#experimental_source strain H37RV
GENETICS
#gene         Rv1165
CLASSIFICATION #superfamily translation elongation factor Tu homology
FEATURE       4-134
SUMMARY       #domain translation elongation factor Tu homology #label
            EF1
            #length 628 #molecular-weight 67580 #checksum 8735
Query Match      59.0%; Score 49; DB 2; Length 628;
Best Local Similarity 54.5%; Pred. No. 1.23e+01;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 363 NVSIRVVVDIGA 373
||:|:| :||:
Qy 73 NASVRTVNI GST 83

Search completed: Thu Jul 8 18:15:59 1999
Job time : 10 secs.

```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:14:48 1999; MasPar time 3.43 Seconds

Tabular output not generated. 115.536 Million cell updates/sec

Title: >US-09-041-236-2
Description: (72-85) from US09041236.pep (6 of 45)
Perfect Score: 83
Sequence: 1 KNASVRTVNIQSTK 14

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 24.529; Variance 25.164; scale 0.975

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	63.9	467	1	ATPX_BACFI	4.56e-01
2	53	63.9	469	1	ATPB_BACFI	4.56e-01
3	52	62.7	499	1	CLK2_MOUSE	7.92e-01
4	52	62.7	499	1	CLK2_HUMAN	7.92e-01
5	52	62.7	570	1	PIL_BACSU	7.92e-01
6	51	61.4	471	1	TNAA_ECOLI	1.36e-00
7	51	61.4	592	1	PIL_ALCEU	1.36e-00
8	51	61.4	738	1	TRFM_HUMAN	1.36e-00
9	50	60.2	572	1	PIL_STAAU	2.33e-00
10	50	60.2	574	1	PIL_STACA	2.33e-00
11	50	60.2	577	1	PIL_LACSK	2.33e-00
12	50	60.2	577	1	PIL_STRSL	2.33e-00
13	50	60.2	577	1	PIL_STRMU	2.33e-00
14	49	59.0	490	1	CLK3_MOUSE	3.96e-00
15	49	59.0	490	1	CLK3_RAT	3.96e-00
16	49	59.0	490	1	CLK3_HUMAN	3.96e-00
17	49	59.0	517	1	DOA_DROME	3.96e-00
18	49	59.0	578	1	PIL_BACST	3.96e-00
19	48	57.8	366	1	HEMN_BACSU	6.65e-00
20	48	57.8	393	1	AAT1_BACSU	6.65e-00
21	48	57.8	409	1	YMXG_BACSU	6.65e-00
22	48	57.8	452	1	TNA2_SYMTH	6.65e-00
23	48	57.8	453	1	TNA1_SYMTH	6.65e-00

24	48	57.8	462	1	TNAA_ENTAE	TRYPTOPHANASE (EC 4.1.1.6.65e-00
25	48	57.8	467	1	TNAA_PROVU	TRYPTOPHANASE (EC 4.1.1.6.65e-00
26	48	57.8	609	1	NEPV_VIBPR	NEUTRAL PROTEASE PRECU 6.65e-00
27	48	57.8	711	1	PTIA_ECOLI	PHOSPHOENOLPYRUVATE-PR 6.65e-00
28	47	56.6	231	1	FLGD_ECOLI	BASAL-BODY ROD MODIFIC 1.11e+01
29	47	56.6	481	1	CLK4_MOUSE	PROTEIN KINASE CLK4 (E 1.11e+01
30	47	56.6	575	1	PT1_ECOLI	PHOSPHOENOLPYRUVATE-PR 1.11e+01
31	47	56.6	575	1	PT1_SALTU	PHOSPHOENOLPYRUVATE-PR 1.11e+01
32	46	55.4	267	1	PANB_SCHPO	PUTATIVE 3-METHYL-2-OX 1.83e+01
33	46	55.4	336	1	G3P2_ANAVA	GLYCERALDEHYDE 3-PHOS 1.83e+01
34	46	55.4	336	1	RECA_MYCPN	RECA PROTEIN. 1.83e+01
35	46	55.4	427	1	AFC2_ARATH	PROTEIN KINASE AFC2 (E 1.83e+01
36	46	55.4	518	1	LEUL_BUCAP	2-ISOPROPYLMALATE SYNT 1.83e+01
37	46	55.4	662	1	COG2_RABIT	72 KD TYPE IV COLLAGEN 1.83e+01
38	46	55.4	748	1	PHLP_ECOLI	PHOSPHOENOLPYRUVATE-PR 1.83e+01
39	46	55.4	838	1	PHSH_SOLFU	ALPHA-GLUCAN PHOSPHORY 1.83e+01
40	45	54.2	80	1	YC15_OENVI	HYPOTHETICAL 9.4 KD PR 2.99e+01
41	45	54.2	87	1	Y212_METJA	HYPOTHETICAL 13.0 KD P 2.99e+01
42	45	54.2	120	1	Y095_MYCTU	HYPOTHETICAL 13.0 KD P 2.99e+01
43	45	54.2	472	1	ATPB_BACME	ATP SYNTHASE BETA CHAI 2.99e+01
44	45	54.2	842	1	PHSH_VICFA	ALPHA-GLUCAN PHOSPHORY 2.99e+01
45	45	54.2	1485	1	CFTR_XENLA	CYSTIC FIBROSIS TRANSM 2.99e+01

ALIGNMENTS

RESULT 1
ID ATPX_BACFI STANDARD; PRT; 467 AA.
AC P25075;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE ATP SYNTHASE BETA CHAIN (EC 3.6.1.34).
GN ATPD
OS BACILLUS FIRMUS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RAB;
RX MEDLINE; 90206802.
RA IVEY D.M., KRULWICH T.A.;
RT "Sequence of the gene encoding the ATP synthase beta subunit from
RT alkaliphilic Bacillus firmus RAB."
RL NUCLEIC ACIDS RES. 18:1296-1296(1990).
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC SUBUNIT.
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A', B AND C.
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31107; G142564; -
CC PIR: S12743; PWBSBF
CC PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
CC PFAM: PF00006; ATP-synt_ab; 1.
CC PFAM: PF00306; ATP-synt_ab_c; 1.
CC HSSP: P07677; 1SKY.
CC -----
CC HYDROLASE; ATP SYNTHESIS; CF(1); ATP-BINDING;
CC HYDROGEN ION TRANSPORT.
CC NP_BIND 156 163
CC FT 250 250 D -> V (IN EMBL DATA LIBRARY).
CC CONFLICT 467 AA; 50712 MW; C80A3515 CRC32;
CC SEQUENCE

Query Match 63.98; Score 53; DB 1; Length 467;
 Best Local Similarity 66.78; Pred. No. 4.56e-01;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 55 DSVRTVMGST 66
 QY 73 NASVRTVNGST 84

RESULT 2

ID ATPB_BACFI STANDARD; PRT; 469 AA.
 AC P22478;
 DT 01-AUG-1991 (REL. 19, CREATED)
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE ATP SYNTHASE BETA CHAIN (EC 3.6.1.34).
 GN ATPD.
 OS BACILLUS FIRMUS.
 OS BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 OC BACILLUS.
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-OF4;
 RX MEDLINE: 92017665.
 RA IVEY D.M., KRULWICH T.A.;
 RT "Organization and nucleotide sequence of the atp genes encoding the
 ATP synthase from alkaliphilic Bacillus firmus OF4.";
 RL MOL. GEN. GENET. 229:292-300(1991).
 CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
 GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
 SUBUNIT.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

EMBL: M60117; GI42551; -;
 PIR: S17726; S17726.
 DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
 DR PFAM: PF00006; ATP-synt_ab; 1.
 DR PFAM: PF00306; ATP-synt_ab_C; 1.
 DR HSP: P07677; ISKY.
 KW HYDROLASE; ATP SYNTHESIS; CF(1); ATP-BINDING;
 FT NP_BIND 157 164 ATP (BY SIMILARITY).
 FT NP_BIND 157 164 HYDROGEN ION TRANSPORT.
 SQ SEQUENCE 469 AA; 50785 MW; B7BEF2F4 CRC32;

Query Match 63.98; Score 53; DB 1; Length 469;
 Best Local Similarity 66.78; Pred. No. 4.56e-01;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 55 DSVRTVMGST 66
 QY 73 NASVRTVNGST 84

RESULT 3

ID CLK2_MOUSE STANDARD; PRT; 499 AA.
 AC Q35491;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE PROTEIN KINASE CLK2 (EC 2.7.1.1-).
 GN CLK2.
 OS HOMO SAPIENS (HUMAN).
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 95082033.
 RA HANES J.J., DER KAMMER H., KLAUDINY J.J., SCHEIT K.H.;
 RT "Characterization by cDNA cloning of two new human protein kinases.
 Evidence by sequence comparison of a new family of mammalian protein
 kinases.";
 RL J. MOL. BIOL. 244:665-672(1994).
 [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 97474796.

GN CLK2.
 OS MUS MUSCULUS (MOUSE).
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 97439710.
 RA NAYLER O., STAMM S., ULLRICH A.;
 RT "Characterization and comparison of four serine- and arginine-rich
 (SR) protein kinases.";
 RL BIOCHEM. J. 326:693-700(1997).
 CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGinine-RICH (SR) PROTEINS
 OF THE SPliceosomal COMPLEX MAY BE A CONStituent OF A NETWORK OF
 REGULATORY MECHANISMS THAT ENABLe SR PROTEINS TO CONtROL RNA
 SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- PTM: AUTOPHOSPHORYLATES ON ALL THREE TYPE OF RESIDUES.
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
 PROTEIN KINASES. BELONGS TO THE LAMMER SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

EMBL: AF033564; G2645852; -;
 DR MGD; MGI:1098669; CLK2.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PFAM: PF00069; pkinase; 1.
 KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
 KW TYROSINE-PROTEIN KINASE; PHOSPHORYLATION; NUCLEAR PROTEIN.
 FT DOMAIN 163 479 PROTEIN KINASE.
 FT NP_BIND 159 177 ATP (BY SIMILARITY).
 FT BINDING 193 193 ATP (BY SIMILARITY).
 FT ACT_SITE 290 290 BY SIMILARITY.
 SQ SEQUENCE 499 AA; 59954 MW; 26703CA2 CRC32;

Query Match 62.78; Score 52; DB 1; Length 499;
 Best Local Similarity 46.2%; Pred. No. 7.92e-01;
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 318 KSTAVRVDFGSA 330
 QY 72 KNASVRTVNGST 84

RESULT 4

ID CLK2_HUMAN STANDARD; PRT; 499 AA.
 AC P49760;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE PROTEIN KINASE CLK2 (EC 2.7.1.1-).
 GN CLK2.
 OS HOMO SAPIENS (HUMAN).
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 95082033.
 RA HANES J.J., DER KAMMER H., KLAUDINY J.J., SCHEIT K.H.;
 RT "Characterization by cDNA cloning of two new human protein kinases.
 Evidence by sequence comparison of a new family of mammalian protein
 kinases.";
 RL J. MOL. BIOL. 244:665-672(1994).
 [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 97474796.


```

RESULT 6
ID TNAI_ECOLI STANDARD; PRT; 471 AA.
AC P00913; P78123;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TRYPTOPHANASE (EC 4.1.99.1) (L-TRYPTOPHAN INDOLE-LYASE) (TNASE).
GN TNAI OR IND.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 82007678.
RA DEELEY M.C., YANOFSKY C.;
RT "Nucleotide sequence of the structural gene for tryptophanase of
RT Escherichia coli K-12."
RL J. BACTERIOL. 147:787-796(1981).
RN [2]
RP SEQUENCE FROM N.A., AND IMPORTANCE OF CYS-298.
RC STRAIN-B/177-A;
RX MEDLINE; 89323226.
RA TOKUSHIGE M., TSUJIMOTO N., ODA T., HONDA T., YUMOTO N., ITO S.,
RA YAMAMOTO M., KIM E.H., HIRAGI Y.;
RT "Role of cysteine residues in tryptophanase for monovalent cation-
RT induced activation."
RL BIOCHIMIE 71:711-720(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 93315143.
RA BURLAND V.D., PLUNKETT G., III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication."
RL GENOMICS 16:551-561(1993).
RN [4]
RP SEQUENCE OF TRYPTIC PEPTIDES.
RC STRAIN-K12;
RX MEDLINE; 72134434.
RA KAGAMIYAMA H., MATSUBARA H., SNELL E.E.;
RT "The chemical structure of tryptophanase from Escherichia coli. 3.
RT Isolation and amino acid sequence of the tryptic peptides."
RL J. BIOL. CHEM. 247:1576-1586(1972).
RN [5]
RP SEQUENCE OF 1-21 FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 86033634.
RA STEWART V., YANOFSKY C.;
RT "Evidence for transcription antitermination control of tryptophanase
RT operon expression in Escherichia coli K-12."
RL J. BACTERIOL. 164:731-740(1985).
RN [6]
RP SEQUENCE OF 463-471 FROM N.A.
RC MEDLINE; 91216998.
RA SARFERO J.P., WOOLFE P.J., GOLLNICK P.D., YANOFSKY C., PITTARD A.J.;
RT "A new family of integral membrane proteins involved in transport of
RT aromatic amino acids in Escherichia coli."
RL J. BACTERIOL. 173:3231-3234(1991).
RN [7]
RP SEQUENCE OF 1-12.
RC STRAIN-K12 / EMG2;
RX MEDLINE; 97443975.
RA LINK A.J., ROBINSON K., CHURCH G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12."
RL ELECTROPHORESIS 18:1259-1313(1997).
RN [8]
RP MUTAGENESIS OF CYS-294 AND CYS-298.
RX MEDLINE; 89278130.
RA PHILLIPS R.S., GOLLNICK P.D.;
RT "Evidence that cysteine 298 is in the active site of tryptophan
indole-lyase."
RL J. BIOL. CHEM. 264:10627-10632(1989).
RN [9]
RP CATALYTIC ACTIVITY: L-TRYPTOPHAN + H(2)O -> INDOLE + PYRUVATE +
RN NH(3) (ALSO CATALYZES THE SYNTHESIS OF TRYPTOPHAN FROM INDOLE
RN AND SERINE).
RN [10]
RP COFACTOR: PYRIDOXAL PHOSPHATE.
RN [11]
RP SUBUNIT: HOMOTETRAMER.
RN [12]
RP SIMILARITY: HIGH, TO TYROSINE PHENOL-LYASE.
RN [13]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN [14]
RP EMBL; K00032; G147997; -.
RN EMBL; X15974; G41936; -.
RN EMBL; M11990; G148001; -.
RN EMBL; M59914; -; NOT_ANNOTATED_CDS.
RN EMBL; L10328; G290556; ALT_INIT.
RN EMBL; AE000448; G1790144; ALT_INIT.
RN F1R; A01136; WZEC.
RN EC02DBASE; G046.5; 6TH EDITION.
RN ECGENE; EG11005; TNAI.
RN PROSITE; PS00853; BETA_ELIM_LYASE; 1.
RN PFAM; PF01212; Beta_elim_lyase; 1.
RN HSSP; P28796; LAX4.
RN TRYPTOPHAN BIOSYNTHESIS; LYASE; PYRIDOXAL PHOSPHATE.
RN BINDING 270 270 PYRIDOXAL PHOSPHATE.
RN MUTAGEN 294 294 C->S: IDENTICAL TO WILD-TYPE.
RN CONFLICT 137 140 DTQ -> TQO (IN REF. 1).
RN CONFLICT 379 380 QA -> TG (IN REF. 1 AND 2).
RN SEQUENCE 471 AA; 52773 MW; 3724A677 CRC32;
Query Match 61.4%; Score 51; DB 1; Length 471;
Best Local Similarity 58.3%; Pred. No. 1.36e+00;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 387 KVAGIRAVEIGS 398
QY 72 KNASVTVNIGS 83
| | | | |
| | | | |
RESULT 7
ID PTL_ALCEU STANDARD; PRT; 592 AA.
AC P23536;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-NOV-1991 (REL. 20, LAST ANNOTATION UPDATE)
DE PHOSPHONOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (EC 2.7.3.9)
DE [PHOSPHOTRANSFERASE SYSTEM, ENZYME I] (PROTEIN I).
DE PHBI.
GN PHBI.
OS ALGALIGENES EUTROPHUS.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; BURKHOLDERIA GROUP;
OC RALSTONIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H16;
RX MEDLINE; 91358378.
RA PRIES A., PRIEFERT H., KRUGER N., STEINBUCHER A.;
RT "Identification and characterization of two Alcaligenes eutrophus
RT gene loci relevant to the poly(beta-hydroxybutyric acid)-leaky
RT phenotype which exhibit homology to ptsH and ptsI of Escherichia
RT coli."
RL J. BACTERIOL. 173:5843-5853(1991).
RN [2]
RP FUNCTION: THIS IS A COMPONENT OF THE PHOSPHONOLPYRUVATE-DEPENDENT
RN SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE
RN -TRANSPORT SYSTEM. ENZYME I TRANSFERS THE PHOSPHORYL GROUP FROM
RN PHOSPHONOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER PROTEIN (HPR).
RN ENZYME I IS COMMON TO ALL PTS.
```

CC -1- CATALYTIC ACTIVITY: PHOSPHOENOLPYRUVATE + PROTEIN HISTIDINE -
 CC PYRUVATE + PROTEIN N-PHOSPHOHISTIDINE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: TO OTHER PEP-UTILIZING ENZYMES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M69036; G141963; -
 CC DR PIR: B38120; B38120
 CC DR PROSITE: PS00370; PEP_ENZYMES_PHOS_SITE; 1.
 CC DR PROSITE: PS00742; PEP_ENZYMES_2; 1.
 CC DR PFAM: PF00391; PEP-utilizers; 1.
 CC DR HSSP: P08839; 2EZA.
 CC KW PHOSPHOTRANSFERASE SYSTEM; TRANSFERASE; KINASE; SUGAR TRANSPORT;
 CC PHOSPHORYLATION.
 CC FT MOD_RES 201 201 PHOSPHORYLATION.
 CC SQ SEQUENCE 592 AA; 65207 MW; 160B299F CRC32;
 CC
 CC Query Match 61.4%; Score 51; DB 1; Length 592;
 CC Best Local Similarity 45.5%; Pred. No. 1.36e+00;
 CC Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 343 TIRTDIGADK 353
 :|:|:|:|:|
 QY 75 SVRTVNIQSTK 85

RESULT 8
 ID TREM_HUMAN STANDARD; PRT: 738 AA.
 AC P08582;
 DT 01-AUG-1988 (REL. 08, CREATED)
 DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE MELANOTRANSFERRIN PRECURSOR (MELANOMA-ASSOCIATED ANTIGEN P97).
 GN MF12 OR MAP97.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MELANOMA;
 RX MEDLINE; 86149285.
 RA ROSE T.M., PLOWMAN G.D., TEPLow D.B., DREYER W.J., HELSTROEM K.E.,
 RA BROWN J.P.;
 RT "Primary structure of the human melanoma-associated antigen p97
 RT (melanotransferrin) deduced from the mRNA sequence."
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:1261-1265(1986).
 RN [2]
 RP GPI-ANCHOR.
 RX MEDLINE; 94132080.
 RA FOOD M.R., ROTHENBERGER S., GABATHULER R., HAIDL I.D., REID G.,
 RA JEFFERIES W.A.;
 RT "Transport and expression in human melanomas of a transferrin-like
 RT glycosylphosphatidylinositol-anchored protein."
 RL J. BIOL. CHEM. 269:3034-3040(1994).
 RN [3]
 RP FUNCTION.
 RX MEDLINE; 96016189.
 RA KENNARD M.L., RICHARDSON D.R., GABATHULER R., PONKA P.,
 RA JEFFERIES W.A.;
 RT "A novel iron uptake mechanism mediated by GPI-anchored human p97."
 RL EMBO J. 14:4178-4186(1995).
 RN [4]
 RP IRON-BINDING.
 RX MEDLINE; 92183868.
 RA BAKER E.N., BAKER H.M., SMITH C.A., STEBBINS M.R., KAHN M.,
 RA HELSTROEM K.E., HELSTROEM J.J.

RT "Human melanotransferrin (p97) has only one functional iron-binding
 RT site".
 RL FEBS LETT. 298:215-218(1992).
 RN [5]
 RP 3D-STRUCTURE MODELLING.
 RX MEDLINE; 92339524.
 RA GARRAT R.C., JHOTI H.;
 RT "A molecular model for the tumour-associated antigen, p97, suggests a
 RT Zn-binding function."
 RL FEBS LETT. 305:55-61(1992).
 CC -1- FUNCTION: INVOLVED IN IRON CELLULAR UPTAKE. SEEMS TO BE
 CC INTERNALIZED AND THEN RECYCLED BACK TO THE CELL MEMBRANE. BINDS A
 CC SINGLE ATOM OF IRON PER SUBUNIT. COULD ALSO BIND ZINC.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
 CC CERTAIN FETAL TISSUES; ALSO FOUND IN LIVER, EPITHELIUM, UMBILICAL
 CC CHORD, PLACENTA AND SWEAT GLAND DUCTS.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M12154; G189518; -
 CC DR EMBL: A00127; G14413; -
 CC DR PIR: A23814; TFHUM.
 CC DR MIM: 155750; -
 CC DR PROSITE: PS00205; TRANSFERRIN_1; 2.
 CC DR PROSITE: PS00206; TRANSFERRIN_2; 2.
 CC DR PROSITE: PS00207; TRANSFERRIN_3; 2.
 CC DR PFAM: PF00405; transferrin; 2.
 CC DR HSSP: P19134; 1TFD.
 CC KW IRON TRANSPORT; GLYCOPROTEIN; TRANSMEMBRANE; DUPLICATION; SIGNAL;
 KW GPI-ANCHOR; MEMBRANE; ZINC.
 FT SIGNAL 1 19
 FT CHAIN 20 ? MELANOTRANSFERRIN.
 FT PROPEP 7 738 REMOVED IN MATURE FORM.
 FT REPEAT 20 361 1.
 FT REPEAT 362 713 2.
 FT DISULFID 26 63 BY SIMILARITY.
 FT DISULFID 36 54 BY SIMILARITY.
 FT DISULFID 130 216 BY SIMILARITY.
 FT DISULFID 172 189 BY SIMILARITY.
 FT DISULFID 186 199 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 FT METAL 78 78 IRON 1 (BY SIMILARITY).
 FT METAL 107 107 IRON 1 (BY SIMILARITY).
 FT METAL 210 210 IRON 1 (BY SIMILARITY).
 FT METAL 279 279 IRON 1 (BY SIMILARITY).
 FT METAL 421 421 IRON 2 (BY SIMILARITY).
 FT METAL 451 451 IRON 2 (BY SIMILARITY).
 FT METAL 536 536 IRON 2 (BY SIMILARITY).
 FT METAL 625 625 IRON 2 (BY SIMILARITY).
 FT BINDING 136 136 ANION (POTENTIAL).
 FT TRANSMEM 714 738 ANCHOR.
 FT CARBOHYD 38 38 POTENTIAL.
 FT CARBOHYD 135 135 POTENTIAL.
 FT CARBOHYD 515 515 POTENTIAL.
 SQ SEQUENCE 738 AA; 80241 MW; F35E5095 CRC32;

Query Match 61.4%; Score 51; DB 1; Length 738;
 Best Local Similarity 54.5%; Pred. No. 1.36e+00;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 674 KDATVRAPVPG 684
 :|:|:|:|:|
 QY 72 KNASVFTVNIQ 82

CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. ENZYME I TRANSFERS THE PHOSPHORYL GROUP FROM
CC PHOSPHOENOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER PROTEIN (HPR).
CC ENZYME I IS COMMON TO ALL PTS.
CC -!- CATALYTIC ACTIVITY: PHOSPHOENOLPYRUVATE + PROTEIN HISTIDINE =
CC PYRUVATE + PROTEIN N-PHOSPHOHISTIDINE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: TO OTHER PEP-UTILIZING ENZYMES.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U82366; G2149152; -
CC PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
CC PROSITE; PS00742; PEP_ENZYMES_2; 1.
CC PFAM; PF00391; PEP-utilizers; 1.
CC HSP; P08839; 2EZA.
CC K W PHOSPHOTRANSFERASE SYSTEM; TRANSFERASE; KINASE; SUGAR TRANSPORT;
KW PHOSPHORYLATION.
FT MOD_RES 190 190 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 574 AA; 63008 MW; 4968F1BF CRC32;

Query Match 60.2%; Score 50; DB 1; Length 574;
Best Local Similarity 50.0%; Pred. No. 2.33e+00;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 328 KPVVVRTMDIGDCK 341
QY 72 KNASVRTVNGSTK 85
| | | | | : | | : | |

RESULT 12
ID PTL_STRLM STANDARD; PRT; 577 AA.
AC P30299;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (EC 2.7.3.9)
DE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I).
GN PTSI.
OS STREPTOCOCCUS SALIVARIUS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25975;
RX MEDLINE; 93051364.
RA GAGNON G., VADEBONCOEUR C., LEVESQUE R.C., FRENETTE M.;
RT "Cloning, sequencing and expression in Escherichia coli of the ptsI
RT gene encoding enzyme I of the phosphoenolpyruvate:sugar
RT phosphotransferase transport system from Streptococcus salivarius";
RL GENE 121:71-78(1992).
RN [2]
RP SEQUENCE OF 1-10 FROM N.A.
RC STRAIN=ATCC 25975;
RX MEDLINE; 94124012;
RA GAGNON G., VADEBONCOEUR C., FRENETTE M.;
RT "Phosphotransferase system of Streptococcus salivarius:
RT characterization of the ptsH gene and its product";
RL GENE 136:27-34(1993).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. ENZYME I TRANSFERS THE PHOSPHORYL GROUP FROM
CC PHOSPHOENOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER PROTEIN (HPR).
CC ENZYME I IS COMMON TO ALL PTS.
CC -!- CATALYTIC ACTIVITY: PHOSPHOENOLPYRUVATE + PROTEIN HISTIDINE =
CC PYRUVATE + PROTEIN N-PHOSPHOHISTIDINE.
CC -!- SUBUNIT: HOMODIMER.
CC -----

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: TO OTHER PEP-UTILIZING ENZYMES.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81756; G153615; -
CC PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
CC PROSITE; PS00742; PEP_ENZYMES_2; 1.
CC PFAM; PF00391; PEP-utilizers; 1.
CC HSP; P08839; 2EZA.
CC K W PHOSPHOTRANSFERASE SYSTEM; TRANSFERASE; KINASE; SUGAR TRANSPORT;
KW PHOSPHORYLATION.
FT MOD_RES 191 191 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 577 AA; 63046 MW; A5D8BD7D CRC32;

Query Match 60.2%; Score 50; DB 1; Length 577;
Best Local Similarity 50.0%; Pred. No. 2.33e+00;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 329 KPVVVRTMDIGDCK 342
QY 72 KNASVRTVNGSTK 85
| | | | | : | | : | |

RESULT 13
ID PTL_STRLM STANDARD; PRT; 577 AA.
AC P45595;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (EC 2.7.3.9)
DE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I).
GN PTSI.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NG5 SEROTYPE C;
RX MEDLINE; 94178918
RA BOYD D.A., CVITKOVITCH D.G., HAMILTON I.R.;
RT "Sequence and expression of the genes for Hpr (ptsH) and enzyme I
RT (ptsI) of the phosphoenolpyruvate-dependent phosphotransferase
RT transport system from Streptococcus mutans";
RL INFECT. IMMUN. 62:1156-1165(1994).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. ENZYME I TRANSFERS THE PHOSPHORYL GROUP FROM
CC PHOSPHOENOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER PROTEIN (HPR).
CC ENZYME I IS COMMON TO ALL PTS.
CC -!- CATALYTIC ACTIVITY: PHOSPHOENOLPYRUVATE + PROTEIN HISTIDINE =
CC PYRUVATE + PROTEIN N-PHOSPHOHISTIDINE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: TO OTHER PEP-UTILIZING ENZYMES.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L15191; G310628; -
CC PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
CC -----

```
DR PROSITE: PS00742; PEP_ENZYMES-2; 1.
DR PFAM: PF00391; PEP-utilizers; 1.
DR HSP: P08839; 2EZA.
KW PHOSPHOTRANSFERASE SYSTEM; TRANSFERASE; KINASE; SUGAR TRANSPORT;
KW PHOSPHORYLATION.
FT MOD_RES 191 191 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 577 AA; 63354 MW; 689F6DB3 CRC32;

Query Match 60.2%; Score 50; DB 1; Length 577;
Best Local Similarity 50.0%; Pred. No. 2.33e+00;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 329 KPVVVTMDIGDK 342
QY 72 KNASVRTVNIQSTK 85

RESULT 14
ID CLK3_MOUSE STANDARD; PRT; 490 AA.
AC O35492;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PROTEIN KINASE CLK3 (EC 2.7.1.-).
GN CLK3.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97439710.
RA NAYLER O., STAMM S., ULLRICH A.;
RT "Characterization and comparison of four serine- and arginine-rich
(RS) protein kinases."
RL BIOCHEM. J. 326:693-700(1997)
CC -!- FUNCTION: PHOSPHORYLATES SERINE- AND ARGinine-RICH (SR) PROTEINS
OF THE SPLICEOSOMAL COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF
REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA
SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- PTM: AUTOPHOSPHORYLATES ON ALL THREE TYPE OF RESIDUES.
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
PROTEIN KINASES. BELONGS TO THE LAMMER SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; X94351; E217388; -
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM: PF00069; Pkinase; 1.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
KW TYROSINE-PROTEIN KINASE; PHOSPHORYLATION; NUCLEAR PROTEIN.
FT DOMAIN 156 472 PROTEIN KINASE.
FT NP_BIND 162 170 ATP (BY SIMILARITY).
FT BINDING 186 186 ATP (BY SIMILARITY).
FT ACT_SITE 283 283 BY SIMILARITY.
SQ SEQUENCE 490 AA; 58501 MW; D9CB161C CRC32;

Query Match 59.0%; Score 49; DB 1; Length 490;
Best Local Similarity 46.2%; Pred. No. 3.96e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 312 KNTSIRVADFGSA 324
QY 72 KNASVRTVNIQST 84

Search completed: Thu Jul 8 18:14:56 1999
Job time : 8 secs.
```

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	83	100.0	393	11	088371	2.59e-07
2	83	100.0	666	4	075326	2.59e-07
3	53	66.3	886	14	Q67652	1.50e+00
4	53	63.9	1847	5	P31495	3.97e+00
5	52	62.7	509	2	Q31149	6.40e+00
6	52	62.7	550	2	Q31692	6.40e+00
7	52	62.7	874	5	Q18109	6.40e+00
8	51	61.4	472	2	Q07674	6.40e+00
9	49	59.0	195	2	Q34412	1.03e+01
10	49	59.0	422	5	Q76584	2.59e+01
11	49	59.0	573	2	Q69251	2.59e+01
12	49	59.0	578	2	Q83018	2.59e+01
13	49	59.0	628	2	Q06563	2.59e+01
14	48	57.8	321	2	Q06013	2.59e+01
15	48	57.8	413	14	Q41627	4.05e+01
16	48	57.8	468	5	Q18358	4.05e+01
17	48	57.8	481	5	Q77353	4.05e+01
18	48	57.8	1408	1	Q27557	4.05e+01
19	47	56.6	95	14	Q74508	6.31e+01
20	47	56.6	95	14	Q74510	6.31e+01

O/5326:	
AC	01-NOV-1998 (TREMBLREL. 08, CREATED)
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE	SEMAPHORIN L.
GS	HOMO SAPIENS (HUMAN).
OS	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC	CATARRHINI; HOMINIDAE; HOMO.
OC	[1]
RN	SEQUENCE FROM N.A.
RP	

RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GERHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
virus";
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030698; G3523115; -.
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 83; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 2.59e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 104 KNASVRTVNIIGSTK 117
QY 72 KNASVRTVNIIGSTK 85

RESULT 3
ID Q67652 PRELIMINARY; PRT; 886 AA.
AC Q67652;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CAPSID PROTEIN.
OS GIARDIAVIRUS SP.
OC VIRUSES; DSRNA VIRUSES; TOTIVIRIDAE; GIARDIAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93391401.
RA WANG A.L., TANG H.M., SHEN K.A., WANG C.C.;
RT "Giardavirus double-stranded RNA genome encodes a capsid polypeptide
RT and a gag-pol-like fusion protein by a translation frameshift.";
RL PROC. NATL. ACAD. SCI. U.S.A. 90:8595-8599(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA WANG A.L.;
RL SUBMITTED (OCT-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA YU D.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; L13218; G1352867; -.
SQ SEQUENCE 886 AA; 98423 MW; A9D17192 CRC32;

Query Match 66.3%; Score 55; DB 14; Length 886;
Best Local Similarity 66.7%; Pred. No. 1.50e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 395 KASINTVVDIGS 406
QY 72 KNASVRTVNIIGS 83

RESULT 4
ID P91495 PRELIMINARY; PRT; 1847 AA.
AC P91495;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILARITY TO RAT INTEGRAL MEMBRANE GLYCOPROTEIN GPI20 PRECURSOR.
GN T23H2.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKIN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WMSLEY P., BRADSHAW H.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U80033; G1703554; -.
SQ SEQUENCE 1847 AA; 199637 MW; B0A25E0F CRC32;

Query Match 63.9%; Score 53; DB 5; Length 1847;
Best Local Similarity 53.8%; Pred. No. 3.97e+00;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 291 KSSSVRALTIGT 303
QY 72 KNASVRTVNIIGST 84

RESULT 5
ID O31149 PRELIMINARY; PRT; 509 AA.
AC O31149;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENZYME I (FRAGMENT).
GN PTSI.
OS LISTERIA MONOCYTOGENES.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC LISTERIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SCOTT A;
RA CHRISTENSEN D.P., BENSON A.K., HUTKINS R.W.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF030824; G2623820; -.
DR PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR PROSITE; PS00742; PEP_ENZYMES_2; 1.
DR PFAM; PF00391; PEP-utilizers; 1.
FT NON_TER 509 509
SQ SEQUENCE 509 AA; 56319 MW; 12A7C5CE CRC32;

Query Match 62.7%; Score 52; DB 2; Length 509;
Best Local Similarity 50.0%; Pred. No. 6.40e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 328 KSVVVRTLDIGDCK 341
QY 72 KNASVRTVNIIGSTK 85

RESULT 6
ID O31692 PRELIMINARY; PRT; 550 AA.
AC O31692;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOTRANSFERASE SYSTEM (PTS) ENZYME I (EC 2.7.3.9).
GN PTSI.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE; 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,

RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA GHM S.Y., GLASER P., GOFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISEPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLSAPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
RA KUBAYASHI Y., KOETTER P., KOENIGSTEIN G., KROGH S., KUMANO M.,
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA PARRO V., POHL T.M., PORTELE D., PORNOLLIK S., PRESCOTT A.M.,
RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SEROR P., SHIN B.S., SOLDI B.,
RA SOROKIN A., TACONSI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMBUUT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
RT "the complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis";
RL NATURE 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: 299111; E1184981;
DR PROSITE: PS00370; PEP-ENZYMES_PHOS_SITE; 1.
DR PROSITE: PS00742; PEP-ENZYMES_2; 1.
DR PFAM: PF00391; PEP-utilizers; 1.
KW TRANSFERASE.
SQ SEQUENCE 550 AA; 60886 MW; 33F19A32 CRC32;

Query Match 62.7%; Score 52; DB 2; Length 550;
Best Local Similarity 50.0%; Pred. No. 6.40e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 307 KSVVVRTLDIGGDK 320
I: |||:|:|:
QY 72 KNASVRTVNIQSTK 85

RESULT 7
ID Q18109 PRELIMINARY; PRT; 874 AA.
AC Q18109;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO THE ACTIN-BINDING.
GN C23F12.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA FULTON B.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U93676; G1049428;
DR PFAM: PF00307; actinin-binding; 1.
DR PFAM: PF00630; Filamin; 2.
SQ SEQUENCE 874 AA; 96987 MW; B0897E75 CRC32;

Query Match 62.7%; Score 52; DB 5; Length 874;
Best Local Similarity 58.3%; Pred. No. 6.40e+00;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 84 REDGVKTVNIGS 95
I: :|:|:|:|:|:
QY 72 KNASVRTVNIQSG 83

RESULT 8
ID O07674 PRELIMINARY; PRT; 472 AA.
AC O07674;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TNAA.
GN TNAA.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98083063.
RA MARTIN K., MORLIN G., SMITH A., NORDYKE A., EISENSTARK A., GOLOMB M.;
RT "The tryptophanase gene cluster of Haemophilus influenzae type b:
RT evidence for horizontal gene transfer";
RL J. BACTERIOL. 180:107-118(1998).
DR EMBL: AF003252; G2209100;
DR PROSITE: PS00853; BETA-ELIM_LYASE; 1.
DR PFAM: PF01212; Beta_elim_lyase; 1.
SQ SEQUENCE 472 AA; 53088 MW; 99DFf409 CRC32;

Query Match 61.4%; Score 51; DB 2; Length 472;
Best Local Similarity 58.3%; Pred. No. 1.03e+01;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 387 KVAGIRAVEIGS 398
I: |:|:|:|:|:
QY 72 KNASVRTVNIQSG 83

RESULT 9
ID Q54412 PRELIMINARY; PRT; 195 AA.
AC Q54412;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE 66 GLYT TRANSFER RNA-GLY(GGA) (FRAGMENT).
OS STREPTOMYCES LIVIDANS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TK64;
RA SCHMIEGER H.;
RL SUBMITTED (MAR-1990) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 121-195 FROM N.A.
RC STRAIN-TK64;
RX MEDLINE; 90326562.
RA SEDLMEIER R., SCHMIEGER H.;
RT "Nucleotide sequences of tRNA genes in Streptomyces lividans 66.";
RN NUCLEIC ACIDS RES. 18:4027-4027(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-TK64;
RX MEDLINE; 94350843.
RA SEDLMEIER R., WERNER T., KIESER H.M., HOPWOOD D.A., SCHMIEGER H.;
RT "rRNA genes of Streptomyces lividans: new sequences and comparison of structure and organization with those of other bacteria.";
RN J. BACTERIOL. 176:5550-5553(1994).
DR EMBL; X52071; G47216; -.
FT NON_TER
SQ SEQUENCE 195 AA; 20350 MW; 8183C8C8 CRC32;

Query Match 59.0%; Score 49; DB 2; Length 195;
Best Local Similarity 53.8%; Pred. No. 2.58e+01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 66 KELGVRRVSGAT 78
I: :|||:|:|:
QY 72 KNASRVTVNIGST 84

RESULT 10
ID 076584 PRELIMINARY; PRT; 422 AA.
AC 076584;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE C16A11.1 PROTEIN.
GN C16A11.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HALLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans";
RN NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA JOHNSON D., BIEWALD T.;
RT "The sequence of C. elegans cosmid C16A11.";
RN SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF077536; G3319382; -.
SQ SEQUENCE 422 AA; 48412 MW; 93CAF826 CRC32;

Query Match 59.0%; Score 49; DB 5; Length 422;
Best Local Similarity 42.9%; Pred. No. 2.58e+01;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 248 KOASLRMVDVGGOR 261
I:|:|:|:|:|:|:
QY 72 KNASRVTVNIGSTK 85

RESULT 11
ID 069251 PRELIMINARY; PRT; 573 AA.
AC 069251;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE ENZYME I.
GN PTSI.
OS BACILLUS MEGATERIUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WH348;
RA WAGNER A., KUESTER E., HILLEN W.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AJ005075; E1285537; -.
DR PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR PROSITE; PS00742; PEP_ENZYMES_2; 1.
SQ SEQUENCE 573 AA; 63658 MW; 5840D124 CRC32;

Query Match 59.0%; Score 49; DB 2; Length 573;
Best Local Similarity 50.0%; Pred. No. 2.58e+01;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 328 KPVVVRTLDIGDK 341
I:|:|:|:|:|:|:
QY 72 KNASRVTVNIGSTK 85

RESULT 12
ID 083018 PRELIMINARY; PRT; 578 AA.
AC 083018;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOTRANSFERASE SYSTEM ENZYME I.
GN PTSI.
OS BACILLUS SP.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S;
RA OHNO M., BEPPU T., UEDA K.;
RT "Bacillus sp. strain S phosphotransferase system enzyme I (ptsI) gene.";
RN SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB016285; D1032916; -.
DR PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR PROSITE; PS00742; PEP_ENZYMES_2; 1.
KW TRANSFERASE; PHOSPHOTRANSFERASE SYSTEM.
SQ SEQUENCE 578 AA; 63451 MW; 1716F96D CRC32;

Query Match 59.0%; Score 49; DB 2; Length 578;
Best Local Similarity 50.0%; Pred. No. 2.58e+01;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```
Db 333 KPVVVRTLDIGGDK 346
| | | | | | | | | |
QY 72 KNASVRTVNIGSTK 85.

RESULT 13
ID 006563 PRELIMINARY; PRT; 628 AA.
AC 006563;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 67.6 KD PROTEIN.
GN RV1165.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE; MYCOBACTERIUM.
OC ACTINOMYCETALES; CORNYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE: 98295987.
RA COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY III C.E., TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RN NATURE 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA PARKHILL J.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: Z95584; E1299888; -
KW HYPOTHETICAL PROTEIN
SQ SEQUENCE 628 AA; 67580 MW; 04C17236 CRC32;

Query Match 59.0%; Score 49; DB 2; Length 628;
Best Local Similarity 54.5%; Pred. No. 2.58e+01;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 363 NVSIRVVVICA 373
| | | | | | | | | |
QY 73 NASVRTVNIGS 83

RESULT 14
ID 006013 PRELIMINARY; PRT; 321 AA.
AC 006013;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COPROPORPHRINOGEN III OXIDASE (FRAGMENT).
GN HEMN.
OS BACILLUS STEAROTHERMOPHILUS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97409967.
RA MOGK A., SCHUMANN W.;
RT "Cloning and sequencing of the hrca gene of Bacillus
RT stearothermophilus.";
RL GENE 194:133-136(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA SCHUMANN W.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: Y09446; E286658; -
FT NON_TER 1 1
```

```
SQ SEQUENCE 321 AA; 37084 MW; EA675D94 CRC32;

Query Match 57.8%; Score 48; DB 2; Length 321;
Best Local Similarity 70.0%; Pred. No. 4.05e+01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 215 VRRVNVIGPIK 224
| | | | | | | | | |
QY 76 VRTVNIGSTK 85

RESULT 15
ID 041627 PRELIMINARY; PRT; 413 AA.
AC 041627;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M20;
RX MEDLINE: 98105804.
RA CONNOR R.I., KORBER B.T.M., GRAHAM B.S., HAHN B.H., HO D.D.,
RA WALKER B.D., NEUMANN A.U., VERMUND S.H., MESTECKY J., JACKSON S.,
RA FENAMORE E., CAO Y., GAO F., KALAMS S., KUNSTMAN K.J., MCDONALD D.,
RA MCWILLIAMS N., TRKOLA A., MOORE J.P., WOJINSKY S.M.;
RT "Immunological and virological analyses of persons infected by human
RT immunodeficiency virus type 1 while participating in trials of
RT recombinant gp120 subunit vaccines.";
RL J. VIROL. 72:1552-1576(1998).
DR EMBL: U84870; G2290134; -
KW ENVELOPE PROTEIN.
FT NON_TER 1 1
FT NON_TER 413 413
SQ SEQUENCE 413 AA; 46189 MW; 60D797AC CRC32;

Query Match 57.8%; Score 48; DB 14; Length 413;
Best Local Similarity 35.7%; Pred. No. 4.05e+01;
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 191 NNNTIRGVHIGPGR 204
| | | | | | | | | |
QY 72 KNASVRTVNIGSTK 85

Search completed: Thu Jul 8 18:15:29 1999
Job time : 13 secs.
```

THIS PAGE BLANK (USPTO)

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on: Thu Jul 8 18:18:56 1999; MasPar time 5.94 Seconds
28.648 Million cell updates/sec
Tabular output not generated.
```

Title: >US-09-041-236-2
Description: (105-112) from US09041236.pep (7 of 45)

Sequence: 1 SEGLLAG 8

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a-genes35

1:part1	2:part2	3:part3	4:part4	5:part5	6:part6	7:part7
8:part8	9:part9	10:part10	11:part11	12:part12	13:part13	
14:part14	15:part15	16:part16	17:part17	18:part18		
19:part19	20:part20	21:part21	22:part22	23:part23		
24:part24	25:part25	26:part26	27:part27	28:part28		
29:part29	30:part30	31:part31	32:part32	33:part33		
34:part34	35:part35	36:part36	37:part37	38:part38		
39:part39						

Statistics: Mean 15.705; Variance 50.779; scale 0.309

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description	Pred. No.
		Match						
1	50	87.7	124	31	W28075		Staphylococcus aureus	8.49e+01
2	50	87.7	245	33	W64230		Human secreted protei	8.49e+01
3	46	80.7	190	26	W23722		Bovine granulocyte pe	2.17e+02
4	46	80.7	929	32	W37259		Rat semaphorin Y	2.17e+02
5	46	80.7	930	32	W37260		Human semaphorin Y	2.17e+02
6	45	78.9	279	2	R10120		Lactococcal tyramidin	2.74e+02
7	45	78.9	319	21	W12377		Regulatory factor inv	2.74e+02
8	45	78.9	748	32	W50143		Cyanobacterial phyto	2.74e+02
9	45	78.9	2861	26	W27227		Human TRIO phosphop	2.74e+02
10	44	77.2	366	25	W18288		Carica papaya ACC syn	3.45e+02
11	44	77.2	479	23	W19771		Beta-1 integrin modul	3.45e+02
12	44	77.2	900	15	R88616		TNF-R p55C-binding p	3.45e+02
13	44	77.2	198	21	W15473		Human PI00 protein	3.45e+02
14	43	75.4	194	36	W72221		HSV-2 strain SB5 Cont	4.33e+02
15	43	75.4	240	9	R45101		Alcaligenes bronchise	4.33e+02
16	43	75.4	240	26	W36101		Mutant arylmalonate d	4.33e+02

Query Match 87.7%; Score 50; DB 31; Length 124;
 Best Local Similarity 71.4%; Pred. No. 8.49e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 19 dglitcg 25
 QY 106 EGLLACG 112

RESULT 2
 ID W64220 standard; Protein; 245 AA.
 AC W64220;
 DT 06-OCT-1998 (first entry)
 DE Human secreted protein from clone CG300_3.
 KW Secreted protein; human adult testes; nutrition; cytokine; stimulant;
 KW cell proliferation; differentiation; immune system; suppressor; ligand;
 KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
 KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
 KW anti-inflammatory.
 OS Homo sapiens.
 PN W09827205-A2.
 PD 25-JUN-1998.
 PF 17-DEC-1997; U23330.
 PR 18-DEC-1997; US-991872.
 PR 18-DEC-1996; US-769192.
 PR 13-JAN-1997; US-783401.
 PA (GENY) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
 PI Rocio LA, Spaulding V, Treacy M;
 DR WPI: 98-362774/31.
 DR N-PSDB: V44294.
 DT New polynucleotides and secreted proteins - obtained from human
 PT foetal brain, human adult testes, human adult brain and human adult
 PT salivary gland cDNA libraries
 PS Claim 14j; Page 68; 110pp; English.
 CC This sequence represents a novel secreted protein from clone CG300_3
 CC isolated from a human adult testes cDNA library. This protein has
 CC applications for nutritional use, cytokine and cell
 CC proliferation/differentiation activity, immune stimulating or
 CC suppressing activity, hematopoiesis regulating activity, tissue growth
 CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
 CC haemostatic and thrombotic activity, receptor/ligand activity,
 CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
 CC tumour inhibition activity and other activities.
 SQ Sequence 245 AA;

Query Match 87.7%; Score 50; DB 33; Length 245;
 Best Local Similarity 75.0%; Pred. No. 8.49e+01;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 183 sqgllncq 190
 QY 105 SEGLLACG 112

RESULT 3
 ID W23722 standard; Protein; 190 AA.
 AC W23722;
 DT 18-FEB-1998 (first entry)
 DE Bovine granulocyte peptide A precursor (antimicrobial BGP-A).
 KW Antimicrobial peptide; antibiotic; antibacterial; antifungal;
 KW fungicide; antiprotzoa; protozoacide; antiviral; virucide;
 KW bovine granulocyte peptide A; BGP-A; preservative; sepsis;
 KW endotoxaemia; cattle.
 OS Bos taurus.
 FH Key Location/Qualifiers
 FT Peptide 1..21 /label= Sig_peptide
 FT Peptide 22..177 /label= Propeptide
 FT Peptide 178..190 /label= Mat_peptide

FT W09729765-A1. /note= "BGP-A antimicrobial peptide (Claim 2)"
 PN 21-AUG-1997.
 PD 13-FEB-1997; U02218.
 PF 16-FEB-1996; US-011834.
 PR (REGC) UNIV CALIFORNIA.
 PA Seilsted ME;
 PI WPI: 97-424753/39.
 DR N-PSDB; T78509.
 DT Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -
 PT useful therapeutically, as preservatives for food, in water
 PT treatment and in agriculture.
 PS Claim 8; Fig 4; 56pp; English.
 CC This protein comprises the precursor (see W23722) of a novel,
 CC claimed antimicrobial peptide from bovine neutrophils, designated
 CC bovine granulocyte peptide A or BGP-A (see W23724). Its amino acid
 CC sequence was deduced from a cDNA clone (see T78509) obtained from
 CC bovine bone marrow. BGP-A and the murine homologue, MGP-A (see
 CC W23725), exhibit activity against Gram-positive and Gram-negative
 CC bacteria, fungi and viruses, specifically Staphylococcus aureus, and C.
 CC Escherichia coli, Candida albicans, Salmonella typhimurium and C.
 CC neofarmans (claimed). They can be used in human or veterinary
 CC medicine (particularly to treat disorders associated with
 CC lipopolysaccharides, e.g. sepsis and endotoxaemia) or as
 CC preservatives in food products or in water supplies (claimed).
 CC They can also be applied to crops to reduce post-harvest spoilage
 CC or expressed in transgenic plants to increase their disease
 CC resistance. They have low immunogenicity. Carboxamidated
 CC analogues of BGP-A and MGP-A may also be used.
 SQ Sequence 190 AA;

Query Match 80.7%; Score 46; DB 26; Length 190;
 Best Local Similarity 62.5%; Pred. No. 2.17e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 142 agsllacg 149
 QY 105 SEGLLACG 112

RESULT 4
 ID W57259 standard; Protein; 929 AA.
 AC W57259;
 DT 02-SEP-1998 (first entry)
 DE Rat semaphorin Y.
 KW Rat; semaphorin Y; nerve extension inhibitor; central nervous system;
 KW peripheral nerve growth.
 OS Rattus norvegicus.
 PN W09811216-A1.
 PD 19-MAR-1998.
 PF 09-SEP-1997; J03167.
 PR 08-AUG-1997; JP-227220.
 PR 11-SEP-1996; JP-263565.
 PA (SUMO) SUMITOMO PHARM CO LTD.
 PI Kikuchi K, Kimura T;
 DR WPI: 98-250958/22.
 DR N-PSDB: V28913, V28914.
 DT DNA encoding human and rat semaphorin Y - an inhibitor of nerve
 PT extension
 PS Claim 1; Page 54-58; 85pp; Japanese.
 CC The present sequence represents rat semaphorin Y, which inhibits nerve
 CC extension. Semaphorin Y genes and proteins may be used to inhibit
 CC peripheral nerve growth. Semaphorin Y antagonists can be used to
 CC accelerate regrowth of the central nervous system.
 SQ Sequence 929 AA;

Query Match 80.7%; Score 46; DB 32; Length 929;
 Best Local Similarity 75.0%; Pred. No. 2.17e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 132 sqtllacg 139
 QY 105 SEGLLACG 112

RESULT 5
ID W57260 standard; Protein: 930 AA.
AC W57260;
DT 02-SEP-1998 (first entry)
DE Human semaphorin Y.
KW Human semaphorin Y; nerve extension inhibitor; central nervous system;
KW peripheral nerve growth.
OS Homo sapiens.
PN W09811216-A1.
PD 19-MAR-1998.
PF 09-SEP-1997; J031167.
PR 08-AUG-1997; JP-227220.
PR 11-SEP-1996; JP-263565.
PA (SUMO) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI; 98-250958/22.
DR N-PSDB; V28915, V28916.
PT DNA encoding human and rat semaphorin Y - an inhibitor of nerve extension
PS Claim 1; Page 65-70; 85pp; Japanese.
CC The present sequence represents human semaphorin Y, which inhibits nerve extension. Semaphorin Y genes and proteins may be used to inhibit peripheral nerve growth. Semaphorin Y antagonists can be used to accelerate regrowth of the central nervous system.
CC Sequence 930 AA;
SQ Sequence 930 AA;

Query Match 80.7%; Score 46; DB 32; Length 930;
Best Local Similarity 75.0%; Pred. No. 2.17e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 133 sqtllacg 140
I:|||||
Qy 105 SEGLLACG 112

RESULT 6
ID R10120 standard; Protein: 279 AA.
AC R10120;
DT 19-MAR-1991 (first entry)
DE Lactococcal thiamidine synthase gene product.
KW TS: vaccine.
OS Streptococcus lactis.
PN EP-406003-A.
PD 02-JAN-1991.
PF 29-JUN-1990; 307140.
PR 30-JUN-1989; IE-002131.
PA (UYCO-) UNIVERSITY COLL COR.
PI O'gara F, Condon S, Ross P;
DR WPI; 91-009312/02.
DR O-PSDB; Q10177.
PT Vector contg. thymidylate synthase as marker gene - provides safe alternative to antibiotic resistance genes for use in agriculture, food, and health prods. and prodn. of attenuated vaccines
PT vaccines
PS Claim 8; Fig 11; 22pp; English.
CC The thymidylate synthetase gene may be used as an alternative to antibiotic marker genes, preferable in that they do not confer antibiotic resistance in potentially pathogenic microbes.
CC Antibiotic resistance in genetically engineered microbes for use in marketing, food processing, health care and attenuated vaccine production.
CC Sequence 279 AA;

Query Match 78.9%; Score 45; DB 2; Length 279;
Best Local Similarity 71.4%; Pred. No. 2.74e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 147 tegllpc 153
I:|||||
Qy 105 SEGLLAC 111

RESULT 7
ID W12377 standard; Protein: 319 AA.
AC W12377;
DT 17-JUN-1997 (first entry)
DE Regulatory factor involved in nitrilase gene expression.
KW Nitrilase; regulatory factor; nitrile; organic acid.
OS Rhodococcus rhodochrous J1 (FERM BP-1478).
PN EP-759474-A2.
PD 26-FEB-1997. 305253.
PF 17-JUL-1996; JP-185626.
PR 21-JUL-1995; JP-185626.
PA (NITT) NITTO CHEM IND CO LTD.
PA (SHIM/) SHIMIZU S.
PI Kobayashi M, Shimizu S;
DR WPI; 97-147520/14.
DR N-PSDB; T63225.
PT Regulatory factor that activates nitrilase gene promoter - and Rhodococcus transformants with increased nitrilase activity, for converting nitrile(s) to organic acids
PS Claim 1; Page 7-8; 13pp; English.
CC A regulatory factor (W12377) that activates the nitrilase gene promoter of Rhodococcus rhodochrous J1 has been identified. The gene (T63225) coding for the regulatory factor is located just downstream of the nitrilase structural gene. The regulatory factor can be used to increase nitrilase expression in Rhodococcus transformants, thus increasing prodn. of an organic acid from its corresp. nitrile. The activity of the regulatory factor is enhanced by the presence of a nitrile, esp. isovaleronitrile.
CC Sequence 319 AA;
SQ Sequence 319 AA;

Query Match 78.9%; Score 45; DB 21; Length 319;
Best Local Similarity 71.4%; Pred. No. 2.74e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 204 sdglllec 210
I:|||||
Qy 105 SEGLLAC 111

RESULT 8
ID W50143 standard; protein: 748 AA.
AC W50143;
DT 28-AUG-1998 (first entry)
DE Cyanobacterial phytochrome Cph1 apoprotein.
KW Phytofluor; fluorescent label; phytochrome; Cph1; cyanobacterium.
OS Synechocystis sp. strain PCC6803.
FH Key Location/Qualifiers
FT Modified_site 538
FT Peptide /note= "undetermined amino acid residue"
FT Peptide 536..544
FT Peptide /note= "histidine kinase transmitter module conserved motif"
FT Peptide 642..653
FT Peptide /note= "histidine kinase transmitter module conserved motif"
FT Peptide 679..687
FT Peptide /note= "histidine kinase transmitter module conserved motif"
FT Peptide 694..698
FT Peptide /note= "histidine kinase transmitter module conserved motif"
FT Peptide 709..714
FT Peptide /note= "histidine kinase transmitter module conserved motif"

PN W09805944-A1.
PD 12-FEB-1998.
PF 01-AUG-1997; U13529.
PR 02-AUG-1996; US-023217.
PA (REGC) UNIV CALIFORNIA.
PI Lagarias JC, Murphy JT;
DR WPI; 98-145711/13.
PT Adducts of apoprotein polypeptide and chromophore as label,

PT particularly for bio-molecules - used as fluorescent markers in
 PT immunoassays, nucleic acid hybridisation, detecting protein-protein
 PT interaction etc., are stable with high molar absorption
 PS Example 4: Fig 10B; 87pp; English.
 CC This polypeptide comprises cyanobacterial phytochrome 1 (Cph1) of
 CC Synechocystis sp. PCC6803. Its amino acid sequence was deduced
 CC from locus sir0473 genomic DNA. Expression of the 748-residue
 CC polypeptide in *E. coli* and incubation with phycocyanobilin yields
 CC an adduct with a red, far-red photo-reversible phytochrome
 CC signature. The invention provides a new class of fluorescent
 CC protein adducts (designated phytofluors) that are generally suitable
 CC for use as fluorescent markers. They comprise a protein component
 CC (an apoprotein) and a bilin chromophore such as phycoerythrobilin.
 CC Preferred apoproteins are obtained from plants, e.g. oats (see
 CC W50144), from green algae, e.g. *Mesotaelium caldarium* (see
 CC W50145), or cyanobacteria such as *Synechocystis*. Truncated
 CC apoproteins consisting of the N-terminal chromophore domain are
 CC especially preferred. Recombinant apoproteins assemble
 CC spontaneously with the bilin chromophore. Claimed compositions
 CC comprise a protein, glycoprotein, antibody or nucleic acid to be
 CC detected linked to the fluorescent adduct. They are used in assays
 CC for detecting the other member of a specific binding pair, e.g.
 CC immunoassay of antigens, immuno-histochemical labelling, as nucleic
 CC acid probes for Southern blotting, for identification of manufactured
 CC products, also to detect protein-protein interactions, including
 CC studies on intracellular protein localisation and identification of
 CC transfected cells. The phytofluors make ideal fluorescent markers
 CC because they have a long wavelength absorption maximum and high
 CC molar absorption coefficient, and are stable to light and pH.
 SQ Sequence 748 AA;

Query Match 78.9%; Score 45; DB 32; Length 748;
 Best Local Similarity 71.4%; Pred. No. 2.74e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 114 sdgllvc 120
 I:|||||
 Qy 105 SEGULLAC 111

RESULT 9
 ID W27227 standard; Protein; 2861 AA.

AC W27227;1998 (first entry)
 DE Human TRIO phosphoprotein.
 KW TRIO; human; proliferative disease; neurodegenerative disease;
 KW oncogene; cytoskeleton; actin; diagnosis; transgenic animal;
 KW antibody; therapy; ss.
 OS Homo sapiens.

FT Key Location/Qualifiers
 FT 252..359
 FT Domain /label= SP-D1
 FT /note= "putative spectrin repeat sequence"
 FT 479..595
 FT Domain /label= SP-D2
 FT /note= "putative spectrin repeat sequence"
 FT 819..925
 FT Domain /label= SP-D3
 FT /note= "putative spectrin repeat sequence"
 FT 1030..1157
 FT Domain /label= SP-D4
 FT /note= "putative spectrin repeat sequence"
 FT 1237..1407
 FT Domain /label= GEF-D1
 FT /note= "GEF domain 1, has rac-1 specificity"
 FT 1435..1534
 FT Domain /label= PSK
 FT /note= "pleckstrin homology domain"
 FT 1914..2085
 FT Domain /label= GEF-D2
 FT /note= "GEF domain 2, has rhoA specificity"
 FT 2113..2214
 FT Domain /label= PSK

FT Domain /note= "pleckstrin homology domain"
 FT 2448..2541
 FT /note= "Ig-like domain"
 FT Domain 2560..2816
 FT /note= "serine/threonine kinase domain"
 FT W09735979-A1.
 PN 02-OCT-1997.
 PD 27-MAR-1997; U052236.
 PR 27-MAR-1996; US-014214.
 PA (DAND) DANA FARBER CANCER INST INC.
 PI Debant A. Serra-Pages C, Streuli M;
 DR WPI; 97-489542/45.
 DR N-PSDB; T85392.
 DR New nucleic acid encoding the TRIO phosphoprotein - used for
 PT diagnosis and treatment of proliferative and neurodegenerative
 PT diseases
 PS Claim 2: Page 93-103; 141pp; English.
 CC This protein sequence comprises the novel multidomain human
 CC phosphoprotein TRIO, which is suggested to be a central organiser of
 CC multiple signaling pathways, to be involved in the activation of
 CC oncogenes such as c-fos, and to induce transformation of cells.
 CC The amino acid sequence was deduced from a composite of cDNA clones
 CC obtained from human WI-38 fibroblast cDNA by interaction trap assay.
 CC TRIO mRNA was expressed in all 8 human tissues examined. Once
 CC isolated, TRIO DNA can be expressed in usual host/vector systems to
 CC produce recombinant TRIO polypeptides, e.g. for antibody production
 CC or screening for modulators of activity. Labelled antibodies that
 CC bind TRIO specifically can be used to detect/quantify TRIO activity,
 CC particularly for diagnosis and phenotyping of neoplastic or
 CC hyperplastic disease. Modulators of TRIO activity (e.g. antisense
 CC nucleic acids, antibodies, peptides or mimetics) can be used to
 CC reorganise the actin cytoskeleton (claimed), e.g. in cases of wound
 CC healing and/or tumour metastasis, to treat an oncogene (claimed),
 CC or more generally to control growth, differentiation, migration
 CC and/or survival of cells, e.g. regulation of the immune response to
 CC infection, treatment of impaired immune response (as in chronic
 CC granulomatous disease), control of apoptosis in cancer therapy, and
 CC treatment of degenerative diseases (e.g. Parkinson's, Alzheimer's
 CC or Huntington's, amyotrophic lateral sclerosis, gastric ulcers,
 CC Wilm's tumour etc.). Transgenic animals can be used as models to
 CC characterise TRIO genes and proteins.
 SQ Sequence 2861 AA;

Query Match 78.9%; Score 45; DB 26; Length 2861;
 Best Local Similarity 62.5%; Pred. No. 2.74e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1648 aeglvpcg 1655
 :|||||
 Qy 105 SEGLLACG 112

RESULT 10

ID W18288 standard; Protein; 366 AA.
 AC W18288;
 DT 27-JAN-1998 (first entry)
 DE Carica papaya ACC synthase capacc2.
 KW ACC synthase; S-adenosyl-L-methionine methylthioadenosine lyase;
 KW capacc2; ethylene biosynthesis; transgenic plant; senescence;
 KW antisense expression system; plant development; fruit ripening;
 KW EC 4.4.1.14; papaya; multigene family.
 OS Carica papaya.
 FT Key Location/Qualifiers
 FT 1..366
 FT Protein /label= capacc2
 FT /note= "Sequence represents 75% of the complete
 ACC synthase protein (i.e. contains no start
 or stop codons)."
 FT W09711166-A1.
 PN 27-MAR-1997.
 PD 20-SEP-1996; AU0591.
 PR 02-MAY-1996; AU-009603.
 PR 20-SEP-1995; AU-005559.

PA (UYOU) UNIV QUEENSLAND.
 PI Botella JR.
 DR WPI: 97-202875/18.
 DR N-PSDB: T76230.
 PT Pineapple, papaya and mango ACC synthase genes - used in gene
 therapy to produce fruits with reduced senescence
 PS Disclosure: Fig 3: 46pp; English.
 CC This sequence represents a novel protein, capacc2, encoded by a member
 of an ACC synthase multigene family found in papaya. The enzyme ACC
 synthase is involved in the pathway for ethylene biosynthesis and the
 rate of endogenous expression of ACC synthase is considered to limit
 CC substantially the rate of ethylene production. Endogenous ethylene is
 often deleterious to crops, especially if some form of mechanical
 CC wounding has occurred and diminishes their post harvest quality and
 CC storage life. Novel ACC synthase genes expressed in transgenic plants
 CC using either sense or antisense expression system may be used to control
 CC the regulation of plant development, in particular fruit ripening,
 CC reducing senescence and thus improving storage life.
 SQ Sequence 366 AA;
 Query Match 77.2%; Score 44; DB 25; Length 366;
 Best Local Similarity 50.0%; Pred. No. 3.45e+02;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 243 ndgvlscg 250
 : : | : | : |
 QY 105 SEGILLACG 112

RESULT 11
 ID W19771 standard; Protein; 479 AA.
 AC W19771;
 DT 22-SEP-1997 (first entry)
 DE Beta-1 integrin modulator B171.
 KW Beta-1 integrin modulator; TRAP2; lymphocyte; inflammation.
 OS Homo sapiens.
 PN W09725618-A1.
 PD 17-JUL-1997.
 PF 03-JAN-1997; U00071.
 PR 05-JAN-1996; US-583318.
 PA (ICOS-) ICOS CORP.
 PI Harris EA, Staunton DE;
 DR WPI: 97-373001/34.
 DR N-PSDB: T72795.
 PT Identification of modulators of integrin binding to proteasome
 PT subunit TRAP2-26S - which are used for regulating immune and
 PT inflammatory responses
 PS Example 1: Page 14-16; 23pp; English.
 CC A polypeptide (W19771) has been identified that interacts with a
 CC beta-1 integrin. Its amino acid sequence was deduced from cDNA
 CC clone B171 (T72795) that was isolated using a two-hybrid system in
 CC yeast to screen for products of a human B cell cDNA library which
 CC interacted with the C-terminus of the cytoplasmic tail of the
 CC beta-1 integrin subunit. The B171-encoded polypeptide shows high
 CC sequence identity to TRAP2, 26S proteasome subunit p97 and 55.11
 CC binding protein. Modulators of beta-1 integrin/TRAP2 interaction
 CC can be used in vivo or in vitro to affect inflammatory processes
 CC involving lymphocytes. Modulators that bind to either the beta-1
 CC integrin or TRAP2 are useful to monitor the level of its binding
 CC partner, either in a body fluid or biopsy sample.
 SQ Sequence 479 AA;
 Query Match 77.2%; Score 44; DB 23; Length 479;
 Best Local Similarity 75.0%; Pred. No. 3.45e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 13 sgallacg 20
 : : | : | : |
 QY 105 SEGILLACG 112

RESULT 12
 ID R88616 standard; Protein; 900 AA.
 AC R88616;
 DT 26-MAR-1996 (first entry)
 DE TNF-R p55IC-binding protein 55.11.
 KW Tumour necrosis factor receptor; TNF-R; p55IC;
 KW intracellular domain binding protein; human immunodeficiency virus;
 KW HIV.
 OS Homo sapiens.
 PN W09531544-A1.
 PD 23-NOV-1995.
 PF 11-MAY-1995; U05854.
 PR 11-MAY-1994; IL-109632.
 PR 02-OCT-1994; IL-111125.
 PA (WEIN/) WEINWURZEL H.
 PA (YEDA) YEDA RES & DEV CO LTD.
 PI Boidin M, Mett I, Varfolomeev E, Wallach D;
 DR WPI: 96-010930/01.
 DR N-PSDB: T03730.
 PT TNF-NGF receptor superfamily intracellular domain-binding proteins -
 PT useful for modulating receptor function, e.g. for treating tumours
 PT or HIV-infected cells
 PS Claim 21; Fig 1d; 96pp; English.
 CC Novel human protein 55.11 (R88616) binds to the intracellular (IC)
 CC domain of tumour necrosis factor receptor (TNF-R) protein p55. It
 CC is obtd. by expression in transformant hosts of 55.11-encoding cDNA
 CC (T03730) isolated from fetal liver and Hela cells. 55.11 can be
 CC used to modulate TNF-R function in the treatment of tumours or
 CC HIV-infected cells, or for the affinity purifn. of other proteins
 CC capable of binding the IC domain of the TNF/NGF receptor
 CC superfamily.
 SQ Sequence 900 AA;
 Query Match 77.2%; Score 44; DB 15; Length 900;
 Best Local Similarity 75.0%; Pred. No. 3.45e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 433 sgallacg 440
 : : | : | : |
 QY 105 SEGILLACG 112

RESULT 13
 ID W15473 standard; Protein; 908 AA.
 AC W15473;
 DT 17-JUN-1997 (first entry)
 DE Human P100 protein.
 KW human; 26S proteasome; P100 protein; protease; yeast; oncogene;
 KW cancer cell; liver; kidney; diagnosis; malignant tumour.
 OS Homo sapiens.
 PN J09075082-A.
 PD 25-MAR-1997.
 PF 11-SEP-1995; 232383.
 PR 11-SEP-1995; JP-232383.
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 DR WPI: 97-239264/22.
 DR N-PSDB: T68421.
 PT Human 26S proteasome component P100 protein - acts to decompose
 PT ubiquitin binding protein, for use in diagnosis and treatment of
 PT cancer
 PS Claim 1: Page 8-11; 15pp; Japanese.
 CC This sequence represents the human 26S proteasome component P100
 CC protein. The proteasome is a multifunctional protease ubiquitous in
 CC eucaryotes ranging from yeasts to human. The 26S proteasome is known to
 CC decompose oncogene products and a significant expression of a proteasome
 CC gene is known to occur in cancer cells in liver, kidney. Elucidation of
 CC the functions of the proteasome will contribute to the diagnosis and
 CC treatment of the conditions such as malignant tumour in which
 CC proteasomes are thought to be involved.
 SQ Sequence 908 AA;
 Query Match 77.2%; Score 44; DB 21; Length 908;
 Best Local Similarity 75.0%; Pred. No. 3.45e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Db 442 sgallacg 449
   I :|||||
QY 105 SEGLLACG 112

RESULT 14
ID W72221 standard; Protein: 194 AA.
AC W72221;
DT 13-JAN-1999 (first entry)
DE HSV-2 strain SB5 Contig ID 15 ORF#63 protein.
KW HSV-2 strain SB5; immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor.
OS Herpes simplex virus type 2.
PN W09820016-A1.
PD 14-MAY-1998.
PF 31-OCT-1997; U20016.
PR 09-JUN-1997; US-049018.
PR 04-NOV-1996; US-030279.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PI Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB,
PI Esser KM, Leary J;
DR WPI: 98-286847/25.
DR N-PSDB; V62176.
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal
PS Claim 10: Page 138: 748pp; English.
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.
CC Based on homology, this sequence is a neural variant mena+ protein.
CC The proteins can be used for the treatment or prevention of disease, to
CC induce an immunological response in a mammal or to identify inhibitors,
CC activators or novel antivirals. Antagonists of the proteins can be used
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC it can also be used to induce an immunological response in a mammal.
SQ Sequence 194 AA:

Query Match 75.4%; Score 43; DB 36; Length 194;
Best Local Similarity 62.5%; Pred. No. 4.33e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 60 sgglmgcg 67
   I :||:|
QY 105 SEGLLACG 112

RESULT 15
ID R45391 standard; Protein: 240 AA.
AC R45391;
DT 11-JUL-1994 (first entry)
DE Alcaligenes bronchisepticus aryl malonate decarboxylase.
KW aryl malonate decarboxylase; AMDase; plasmid pAMD100; prokaryote.
OS Alcaligenes bronchisepticus KUI201.
FH Key Location/Qualifiers
FT misc_difference 28
FT /note= "corresponds to TAT codon"
FT misc_difference 233
FT /note= "corresponds to TAC codon"
PN J05328976-A.
PD 14-DEC-1993.
PF 29-MAY-1992; 137364.
PR 29-MAY-1992; JP-137364.
PA (HITB ) HITACHI CHEM CO LTD.
PA (OTAH/) OTA H.
DR WPI: 94-021924/03.
DR N-PSDB; Q54496.
PT New DNA fragment - contg. base sequence coding aryl-malonate
PT decarboxylase
PS Claim 1; Page 2; 12pp; Japanese.
CC Genomic DNA was isolated from A.bronchisepticus KUI201 and
CC fragmented by restriction enzyme digestion. Fragments were ligated

```

```

CC to pUC19 and recombinant plasmids were used to transform competent
CC E.coli DH5alpha-MCR cells. The plasmid pAMD100 was isolated from a
CC transformant which expressed aryl malonate decarboxylase activity.
CC R45391 is the amino acid sequence of the AMDase encoded by the
CC plasmid insert.
SQ Sequence 240 AA:

```

```

Query Match 75.4%; Score 43; DB 9; Length 240;
Best Local Similarity 66.7%; Pred. No. 4.33e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

```

```

Db 181 sdgillscg 189
   I :||:|
QY 105 SEG-LLACG 112

```

```

Search completed: Thu Jul 8 18:19:19 1999
Job time : 23 secs.

```

MPSRCH_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:18:21 1999; MasPar time 3.13 Seconds
Tabular output not generated.
Title: >US-09-041-236-2
Description: (105-112) from US09041236.pep (7 of 45)
Perfect Score: 57
Sequence: 1 SEGLLACG 8
Scoring table: PAM 150
Gap 11
Searched: 122810 seqs, 40068593 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4
Statistics: Mean 21.875; Variance 25.805; scale 0.848
pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Query Match Length DB ID Description Pred. No.

1 54 94.7 653 2 T03102 semaphorin homolog A3 2.06e-01
2 50 87.7 286 2 I80309 sepB protein - Escherichia coli 1.66e+00
3 50 87.7 335 2 S37304 spaL protein - Salmonella 1.66e+00
4 50 87.7 434 2 D71823 flagellum-specific AT 1.66e+00
5 50 87.7 434 2 D64697 flagellar export prot 1.66e+00
6 50 87.7 439 2 A55320 yscN protein - Yersinia 1.66e+00
7 47 82.5 327 2 T01038 myb-related transcrip 7.42e+00
8 47 82.5 389 2 S62957 hypothetical protein 7.42e+00
9 47 82.5 467 2 T02238 glucosyl transferase, 7.42e+00
10 47 82.5 587 2 B70884 hypothetical protein 7.42e+00
11 46 80.7 279 2 A43797 thymidylate synthase 1.20e+01
12 46 80.7 392 2 H65004 hypothetical protein 1.20e+01
13 46 80.7 406 2 D69878 pantothenate metaboli 1.20e+01
14 45 78.9 161 2 T01777 dynein-like protein 7 1.94e+01
15 45 78.9 277 2 G71456 hypothetical protein 1.94e+01
16 45 78.9 301 2 D76023 dihydrodipicolinate s 1.94e+01
17 45 78.9 319 2 JG6117 transcription factor 1.94e+01
18 45 78.9 748 1 S74389 phytochrome phy - Syn 3.11e+01
19 44 77.2 83 2 S70360 exob protein homolog 3.11e+01
20 44 77.2 180 2 S47662 transcription elongat 3.11e+01
21 44 77.2 180 2 S47662 transcription elongat 3.11e+01
22 44 77.2 255 1 XVEG1 trNA (guanine-N1)-me 3.11e+01
23 44 77.2 331 2 G70942 hypothetical protein 3.11e+01

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd
MPSRCH_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:18:21 1999; MasPar time 3.13 Seconds
Tabular output not generated.
Title: >US-09-041-236-2
Description: (105-112) from US09041236.pep (7 of 45)
Perfect Score: 57
Sequence: 1 SEGLLACG 8
Scoring table: PAM 150
Gap 11
Searched: 122810 seqs, 40068593 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4
Statistics: Mean 21.875; Variance 25.805; scale 0.848
pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Query Match Length DB ID Description Pred. No.

1 54 94.7 653 2 T03102 semaphorin homolog A3 2.06e-01
2 50 87.7 286 2 I80309 sepB protein - Escherichia coli 1.66e+00
3 50 87.7 335 2 S37304 spaL protein - Salmonella 1.66e+00
4 50 87.7 434 2 D71823 flagellum-specific AT 1.66e+00
5 50 87.7 434 2 D64697 flagellar export prot 1.66e+00
6 50 87.7 439 2 A55320 yscN protein - Yersinia 1.66e+00
7 47 82.5 327 2 T01038 myb-related transcrip 7.42e+00
8 47 82.5 389 2 S62957 hypothetical protein 7.42e+00
9 47 82.5 467 2 T02238 glucosyl transferase, 7.42e+00
10 47 82.5 587 2 B70884 hypothetical protein 7.42e+00
11 46 80.7 279 2 A43797 thymidylate synthase 1.20e+01
12 46 80.7 392 2 H65004 hypothetical protein 1.20e+01
13 46 80.7 406 2 D69878 pantothenate metaboli 1.20e+01
14 45 78.9 161 2 T01777 dynein-like protein 7 1.94e+01
15 45 78.9 277 2 G71456 hypothetical protein 1.94e+01
16 45 78.9 301 2 D76023 dihydrodipicolinate s 1.94e+01
17 45 78.9 319 2 JG6117 transcription factor 1.94e+01
18 45 78.9 748 1 S74389 phytochrome phy - Syn 3.11e+01
19 44 77.2 83 2 S70360 exob protein homolog 3.11e+01
20 44 77.2 180 2 S47662 transcription elongat 3.11e+01
21 44 77.2 180 2 S47662 transcription elongat 3.11e+01
22 44 77.2 255 1 XVEG1 trNA (guanine-N1)-me 3.11e+01
23 44 77.2 331 2 G70942 hypothetical protein 3.11e+01

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd
MPSRCH_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:18:21 1999; MasPar time 3.13 Seconds
Tabular output not generated.
Title: >US-09-041-236-2
Description: (105-112) from US09041236.pep (7 of 45)
Perfect Score: 57
Sequence: 1 SEGLLACG 8
Scoring table: PAM 150
Gap 11
Searched: 122810 seqs, 40068593 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4
Statistics: Mean 21.875; Variance 25.805; scale 0.848
pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Query Match Length DB ID Description Pred. No.

1 54 94.7 653 2 T03102 semaphorin homolog A3 2.06e-01
2 50 87.7 286 2 I80309 sepB protein - Escherichia coli 1.66e+00
3 50 87.7 335 2 S37304 spaL protein - Salmonella 1.66e+00
4 50 87.7 434 2 D71823 flagellum-specific AT 1.66e+00
5 50 87.7 434 2 D64697 flagellar export prot 1.66e+00
6 50 87.7 439 2 A55320 yscN protein - Yersinia 1.66e+00
7 47 82.5 327 2 T01038 myb-related transcrip 7.42e+00
8 47 82.5 389 2 S62957 hypothetical protein 7.42e+00
9 47 82.5 467 2 T02238 glucosyl transferase, 7.42e+00
10 47 82.5 587 2 B70884 hypothetical protein 7.42e+00
11 46 80.7 279 2 A43797 thymidylate synthase 1.20e+01
12 46 80.7 392 2 H65004 hypothetical protein 1.20e+01
13 46 80.7 406 2 D69878 pantothenate metaboli 1.20e+01
14 45 78.9 161 2 T01777 dynein-like protein 7 1.94e+01
15 45 78.9 277 2 G71456 hypothetical protein 1.94e+01
16 45 78.9 301 2 D76023 dihydrodipicolinate s 1.94e+01
17 45 78.9 319 2 JG6117 transcription factor 1.94e+01
18 45 78.9 748 1 S74389 phytochrome phy - Syn 3.11e+01
19 44 77.2 83 2 S70360 exob protein homolog 3.11e+01
20 44 77.2 180 2 S47662 transcription elongat 3.11e+01
21 44 77.2 180 2 S47662 transcription elongat 3.11e+01
22 44 77.2 255 1 XVEG1 trNA (guanine-N1)-me 3.11e+01
23 44 77.2 331 2 G70942 hypothetical protein 3.11e+01

24 44 77.2 348 2 A71153 hypothetical protein 3.11e+01
25 44 77.2 361 2 G70862 probable adhe2 protei 3.11e+01
26 44 77.2 430 2 C42284 Spa47 protein - Shige 3.11e+01
27 44 77.2 591 2 G01586 DNA helicase II - Vib 3.11e+01
28 44 77.2 611 2 S57527 55.11 protein homolog 3.11e+01
29 44 77.2 900 2 S66264 meiotic recombination 3.11e+01
30 44 77.2 926 2 S58936 hypothetical protein 3.11e+01
31 44 77.2 1398 2 H71606 hypothetical protein 3.11e+01
32 43 75.4 156 2 S76395 hypothetical protein 4.94e+01
33 43 75.4 240 2 A48374 arylmalonate decarbox 4.94e+01
34 43 75.4 240 2 JS0754 trNA (guanine-N1)-me 4.94e+01
35 43 75.4 246 2 C64054 trNA (guanine-N1)-me 4.94e+01
36 43 75.4 255 2 S37175 hypothetical protein 4.94e+01
37 43 75.4 269 2 C70710 hypothetical protein 4.94e+01
38 43 75.4 464 2 S50592 probable oxaloacetate 4.94e+01
39 43 75.4 469 2 C71373 hypothetical protein 4.94e+01
40 43 75.4 607 2 S01939 hypothetical protein 4.94e+01
41 43 75.4 618 2 S33044 hypothetical protein 4.94e+01
42 43 75.4 634 2 S56817 probable succinate de 4.94e+01
43 43 75.4 777 1 WMA041 late 100K protein - h 4.94e+01
44 43 75.4 1091 2 A44147 calcium channel prote 4.94e+01
45 43 75.4 1308 2 A47253 epidermal growth fact 4.94e+01

ALIGNMENTS

RESULT 1
ENTRY T03102 #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS T03102 preliminary; translated from GB/EMBL/DBJ
REFERENCE 214840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-653 #label ENS
#cross-references EMBL:AF005370; NID:q2337967; PID:q2337970
SUMMARY #length 653 #molecular-weight 73645 #checksum 5501

Query Match 94.7%; Score 54; DB 2; Length 653;
Best Local Similarity 75.0%; Pred. No. 2.06e-01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 159 TDGLLACG 166
Qy 105 SEGLLACG 112
RESULT 2
ENTRY I80309 #type complete
TITLE sepB protein - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 07-Jun-1996
ACCESSIONS I80309
REFERENCE I59412
#authors Jarvis, K.G.; Giron, J.A.; Jerse, A.E.; McDaniel, T.K.;
Donnenberg, M.S.; Kaper, J.B.
#journal Proc. Natl. Acad. Sci. U.S.A. (1995) 92:7996-8000
#title Enteropathogenic Escherichia coli contains a putative type III secretion system necessary for the export of proteins involved in attaching and effacing lesion formation.
#cross-references MUID:95372406
#accession I80309 preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-286 #label RES
#cross-references EMBL:Z49933; NID:g887432; PID:g887434

```

SUMMARY      #length 286  #molecular-weight 31467  #checksum 7372

Query Match      87.7%; Score 50; DB 2; Length 286;
Best Local Similarity 71.4%; Pred. No. 1.66e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 162 DGLTTCG 168
Qy 106 EGLLAGC 112
      :|||:|

RESULT 3
ENTRY   S37304      #type fragment
TITLE   spaL protein - Salmonella typhimurium (fragment)
ORGANISM #formal_name Salmonella typhimurium
DATE     31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
18-Sep-1998

ACCESSIONS
REFERENCE S37304
#authors  Groisman, E.A.; Ochman, H.
#journal  EMBO J. (1993) 12:3779-3787
#title    Cognate gene clusters govern invasion of host epithelial
          cells by Salmonella typhimurium and Shigella flexneri.
#cross-references MUID:94008985
#accession S37304      nucleic acid sequence not shown
#status     nucleic acid sequence not shown
#molecule_type DNA
#residues   1-335 #label GRO
#cross-references EMBL:X73525; NID:g404287

GENETICS
#gene       spaL
CLASSIFICATION #superfamily H+-transporting ATP synthase alpha chain;
               H+-transporting ATP synthase alpha chain homology
63-70
86-254
#region nucleotide-binding motif A (P-loop)\
#domain H+-transporting ATP synthase alpha chain
homology #label ATP
#length 335 #checksum 2135

SUMMARY      87.7%; Score 50; DB 2; Length 335;
Best Local Similarity 71.4%; Pred. No. 1.66e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 48 DGLTTCG 54
Qy 106 EGLLAGC 112
      :|||:|

RESULT 4
ENTRY   D71823      #type complete
TITLE   flagellum-specific ATP synthase - Helicobacter pylori (strain
          J99)
ORGANISM #formal_name Helicobacter pylori
#variety  strain J99
DATE     12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
05-Mar-1999

ACCESSIONS
REFERENCE D71823
#authors  Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
          Dolg, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
          B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
          Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
          Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
          G.F.; Trust, T.J.
          Nature (1999) 397:176-180
#journal  Genomic sequence comparison of two unrelated isolates of the
          human gastric pathogen Helicobacter pylori.
#title
#cross-references MUID:99120557
#accession D71823
#status     preliminary
#molecule_type DNA
#residues   1-434 #label ARN
#cross-references GB:AE001554; GB:AE001439; NID:g4155914; PID:g4155920

```

```

GENETICS      ##experimental_source strain J99
#gene         fliI
CLASSIFICATION #superfamily H+-transporting ATP synthase alpha chain;
               H+-transporting ATP synthase alpha chain homology
SUMMARY      #length 434 #molecular-weight 47703 #checksum 4558

Query Match      87.7%; Score 50; DB 2; Length 434;
Best Local Similarity 71.4%; Pred. No. 1.66e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 149 DGLTTCG 155
Qy 106 EGLLAGC 112
      :|||:|

RESULT 5
ENTRY   D64697      #type complete
TITLE   flagellar export protein ATP synthase - Helicobacter pylori
          (strain 26695)
ORGANISM #formal_name Helicobacter pylori
DATE     09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
21-Aug-1998

ACCESSIONS
REFERENCE D64697
#authors  Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
          Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
          H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
          J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
          Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
          McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
          Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
          Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
          Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
          W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
          C.M.; Venter, J.C.
          Nature (1997) 388:539-547
#journal  The complete genome sequence of the gastric pathogen
          Helicobacter pylori.
#title
#cross-references MUID:97394467
#accession D64697
#status     preliminary; nucleic acid sequence not shown;
          translation not shown
#molecule_type DNA
#residues     1-434 #label TOM
#cross-references GB:AE000642; GB:AE000511; NID:g2314587; PID:g2314594;
          TIGR:HP1420

CLASSIFICATION #superfamily H+-transporting ATP synthase alpha chain;
               H+-transporting ATP synthase alpha chain homology
KEYWORDS
FEATURE      P-loop
164-171
SUMMARY      #region nucleotide-binding motif A (P-loop)
               #length 434 #molecular-weight 47636 #checksum 4635

Query Match      87.7%; Score 50; DB 2; Length 434;
Best Local Similarity 71.4%; Pred. No. 1.66e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 149 DGLTTCG 155
Qy 106 EGLLAGC 112
      :|||:|

RESULT 6
ENTRY   A55520      #type complete
TITLE   yscN protein - Yersinia enterocolitica
ORGANISM #formal_name Yersinia enterocolitica
DATE     25-Aug-1995 #sequence_revision 25-Aug-1995 #text_change
21-Aug-1998

ACCESSIONS
REFERENCE A55520; H35392
#authors  Woestyn, S.; Allaoui, A.; Wattiau, P.; Cornelis, G.R.
          J. Bacteriol. (1994) 176:1561-1569
#journal

```


#title yscN, the putative energizer of the Yersinia Yop secretion machinery.
 #cross-references MUID:94179088
 #accession A5520
 ##status preliminary
 ##molecule_type DNA
 ##residues 1-439 #label WOE
 ##cross-references GB:U02499; NID:g437201; PID:g437202
 REFERENCE A35392
 #authors Viitanen, A.M.; Toivanen, P.; Skurnik, M.
 #journal J. Bacteriol. (1990) 172:3152-3162
 #title The lcrE gene is part of an operon in the lcr region of Yersinia enterocolitica O:3.
 #cross-references MUID:90264308
 #accession H35392
 ##status preliminary
 ##molecule_type DNA
 ##residues 94-150 #label VII
 ##cross-references GB:M32097

GENETICS
 #gene yscN
 CLASSIFICATION #superfamily H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase alpha chain homology
 KEYWORDS P-loop
 FEATURE 169-176 #region nucleotide-binding motif A (P-loop)\
 192-360 #domain H+-transporting ATP synthase alpha chain homology #label ATP
 SUMMARY #length 439 #molecular-weight 47782 #checksum 4600
 Query Match 87.7%; Score 50; DB 2; Length 439;
 Best Local Similarity 71.4%; Pred. No. 1.56e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 154 DGLITCG 160
 :|||||
 QY 106 EGLLAGC 112

RESULT 7
 ENTRY T01038 #type complete
 TITLE myb-related transcription factor homolog YUP8H12R.21 - Arabidopsis thaliana
 ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear cress
 DATE 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 05-Feb-1999
 ACCESSIONS T01038
 REFERENCE 214227
 #authors Theologis, A.; Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; Kwan, A.; Toriumi, M.; Yu, G.; Oji, O.; Araujo, R.; Chung, E.; Dewar, K.; Dietrich, F.; Ecker, J.R.; Marziani, A.; Oefner, P.; Davis, R.W.
 #submission submitted to the EMBL Data Library; May 1998
 #description Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.
 #accession T01038
 ##status preliminary; translated from GB/EMBL/DBJ
 ##molecule_type DNA
 ##residues 1-327 #label THE
 ##cross-references EMBL:AC002986; NID:g2494106; PID:g3152563

GENETICS
 #map_position 1
 #introns 123/2
 #note YUP8H12R.21
 SUMMARY #length 327 #molecular-weight 37643 #checksum 9701
 Query Match 82.5%; Score 47; DB 2; Length 327;
 Best Local Similarity 75.0%; Pred. No. 7.42e+00;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 78 NSGLRCG 85
 :|||||
 QY 105 SEGLLAGC 112

RESULT 8

ENTRY S62957 #type complete
 TITLE hypothetical protein YNL035c - yeast (Saccharomyces cerevisiae)
 ALTERNATE_NAMES hypothetical protein N2730
 ORGANISM #formal_name Saccharomyces cerevisiae
 DATE 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 09-Apr-1998
 ACCESSIONS S62957
 REFERENCE S62944
 #authors Duesterhoeft, A.; Floeth, M.; Fritz, C.; Heuss-Neitzel, D.; Hilbert, H.; Mostl, D.
 #submission submitted to the Protein Sequence Database, April 1996
 #accession S62957
 ##molecule_type DNA
 ##residues 1-389 #label DUE
 ##cross-references EMBL:Z71311; NID:g1301877; PID:e239670; PID:g1301880; MIPS:YNL035c

GENETICS
 ##experimental_source strain S288C

CLASSIFICATION #superfamily WD repeat homology

FEATURE 144-178 #domain WD repeat homology #label WD1

SUMMARY #length 389 #molecular-weight 43848 #checksum 9365
 Query Match 82.5%; Score 47; DB 2; Length 389;
 Best Local Similarity 100.0%; Pred. No. 7.42e+00;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 113 GLLACG 118
 :|||||
 QY 107 GLLACG 112

RESULT 9

ENTRY T02238 #type complete
 TITLE glucosyl transferase, jasmonate-induced - common tobacco
 ORGANISM #formal_name Nicotiana tabacum #common_name common tobacco
 DATE 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Mar-1999
 ACCESSIONS T02238
 REFERENCE 214633
 #authors Kojima, H.; Hashizume, K.; Imanishi, S.; Nakamura, K.
 #submission submitted to the EMBL Data Library, January 1997
 #description Jasmonate-induced potential glucosyltransferase from tobacco suspension cell.

GENETICS
 #accession T02238
 ##status preliminary; translated from GB/EMBL/DBJ
 ##molecule_type mRNA
 ##residues 1-467 #label KOJ
 ##cross-references EMBL:AB000623; NID:dl094897; PID:dl019901
 ##experimental_source strain BY-2

GENETICS
 #gene JIGT
 SUMMARY #length 467 #molecular-weight 52989 #checksum 1164
 Query Match 82.5%; Score 47; DB 2; Length 467;
 Best Local Similarity 75.0%; Pred. No. 7.42e+00;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 298 SOGLKCG 305
 :|||||
 QY 105 SEGLLAGC 112

RESULT 10

ENTRY B70884 #type complete
 TITLE hypothetical protein RV2787 - Mycobacterium tuberculosis (strain H37RV)
 ORGANISM #formal_name Mycobacterium tuberculosis

```

DATE      17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
ACCESSIONS B70884
REFERENCE   A70500
#authors    Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
            C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
            III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
            Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
            Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
            Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
            Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
            Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
            Skellerton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
            Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal    Nature (1998) 393:537-544
#title      Deciphering the biology of Mycobacterium tuberculosis from
            the complete genome sequence.
#cross-references MUID:98295987
#accession   B70884
#status      preliminary: nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues    1-587 #label COL
#cross-references GB:AL008967; GB:AL123456; NID:g3261491; PID:ell73917;
            PID:g2624309
#experimental_source strain H37Rv
GENETICS    Rv2787
#gene        #length 587 #molecular-weight 63848 #checksum 841
SUMMARY
Query Match      82.5%; Score 47; DB 2; Length 587;
Best Local Similarity 100.0%; Pred. No. 7.42e+00;
Matches          0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      46 GLLACG 51
        |||||
Qy     107 GLLACG 112

RESULT      11
ENTRY       A43797
TITLE       thymidylate synthase (EC 2.1.1.45) - Lactococcus lactis
            subsp. lactis
ORGANISM    #formal_name Lactococcus lactis subsp. lactis
DATE        12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
            08-Sep-1997
ACCESSIONS  A43797
REFERENCE    Ross, P.; O'Gara, F.; Condon, S.
            Appl. Environ. Microbiol. (1990) 56:2156-2163
            Cloning and characterization of the thymidylate synthase gene
            from Lactococcus lactis subsp. lactis.
#cross-references MUID:90358505
#accession   A43797
#molecule_type DNA
#residues    1-279 #label ROS
#cross-references GB:M33770; NID:g149512; PID:g149513
GENETICS    thya
#gene        #domain thymidylate synthase homology #label TDS
            DNA biosynthesis: methyitransferase
KEYWORDS     #superfamily thymidylate synthase; thymidylate synthase
            homology
FEATURE      DNA biosynthesis: methyitransferase
5-279        #length 279 #molecular-weight 32577 #checksum 5380
SUMMARY
Query Match      80.7%; Score 46; DB 2; Length 279;
Best Local Similarity 62.5%; Pred. No. 1.20e+01;
Matches          3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      147 TEGLLPCA 154
        |||||
Qy     105 SEGLLACG 112

```

```

RESULT      12
ENTRY       H65004
TITLE       #type complete
            hypothetical protein b2322 - Escherichia coli (strain K-12)
ORGANISM    #formal_name Escherichia coli
DATE        12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
            21-Nov-1997
ACCESSIONS  H65004
REFERENCE    A64720
            Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
            Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
            Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
            Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
            Y.
            Science (1997) 277:1453-1462
#journal     Science (1997) 277:1453-1462
#title       The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession   H65004
#status      preliminary: nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues    1-392 #label BLAT
#cross-references GB:AE000321; GB:U00096; NID:gl788659; PID:gl788662;
            UMGp:b2322
#experimental_source strain K-12, substrain MG1655
CLASSIFICATION #superfamily hypothetical protein b2322
SUMMARY        #length 392 #molecular-weight 40465 #checksum 9841

Query Match      80.7%; Score 46; DB 2; Length 392;
Best Local Similarity 71.4%; Pred. No. 1.20e+01;
Matches          5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db      86 QGMLACG 92
        |||||
Qy     106 EGLLACG 112

RESULT      13
ENTRY       D69878
TITLE       #type complete
            pantothenate metabolism flavoprotein homolog yloI - Bacillus
            subtilis
ORGANISM    #formal_name Bacillus subtilis
DATE        05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
            24-Sep-1998
ACCESSIONS  D69878
REFERENCE    Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
            Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
            Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
            A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
            Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
            Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
            Daniel, R.A.; Denizot, F.; Devine, K.M.; Duisterhoft, A.;
            Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
            Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
            M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Galleron, N.; Ghim,
            S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
            Guiseppe, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
            C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
            Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karanata, D.;
            Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
            Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
            Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
            Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
            Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
            M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
            M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
            V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
            A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
            Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
            Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
            Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;

```

Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

#cross-references MUID:98044033

#accession D69878

#status preliminary; nucleic acid sequence not shown;

translation not shown

#molecule_type DNA

##residues 1-406 ##label KUN

##cross-references GB:D299112; GB:AL009126; NID:g2633902; PID:el185161;

PID:g2633942

##experimental_source strain 168

GENETICS

#gene yloI

CLASSIFICATION #superfamily pantothenate metabolism flavoprotein dfp

SUMMARY #length 406 #molecular-weight 43977 #checksum 915

Query Match 80.7%; Score 46; DB 2; Length 406;

Best Local Similarity 87.5%; Pred. NO. 1.20e+01;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 151 SEGYLACG 158

Qy 105 SEGLLACG 112

RESULT 14

ENTRY

TITLE dynein-like protein 7 - rat (fragment)

ORGANISM #formal_name Rattus norvegicus #common_name Norway rat

DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change

28-Feb-1997

ACCESSIONS I70177

REFERENCE I55515

authors Tanaka, Y.; Zhang, Z.; Hirokawa, N.

#journal J. Cell Sci. (1995) 108:1883-1893

#title Identification and molecular evolution of new dynein-like

protein sequences in rat brain.

#cross-references MUID:95386588

#accession I70177

#status preliminary; translated from GB/EMBL/DBJ

#molecule_type mRNA

##residues 1-161 ##label RES

##cross-references GB:D26498; NID:g871910; PID:g871911

GENETICS DLP7

#gene

SUMMARY #length 161 #checksum 5456

Query Match

Best Local Similarity 78.9%; Score 45; DB 2; Length 161;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 34 GLLSCG 39

Qy 107 GLLACG 112

RESULT 15

ENTRY

TITLE hypothetical protein PH0309 - *Pyrococcus horikoshii*

ORGANISM #formal_name *Pyrococcus horikoshii*

DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change

14-Aug-1998

ACCESSIONS G71456

REFERENCE

#authors

A71000

Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

#journal

#title

DNA Res. (1998) 5:55-76
Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.

#cross-references MUID:98344137

#accession G71456

#status preliminary; nucleic acid sequence not shown;

translation not shown

#molecule_type DNA

##residues 1-277 ##label KAW

##cross-references GB:AP000001; NID:g3236128; PID:d1030325; PID:g3256699

##experimental_source strain OT3

##note this accession replaces an interim accession for a sequence replaced by GenBank

GENETICS

#gene

SUMMARY

PH0309 #length 277 #molecular-weight 32345 #checksum 4996

Query Match

Best Local Similarity 78.9%; Score 45; DB 2; Length 277;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

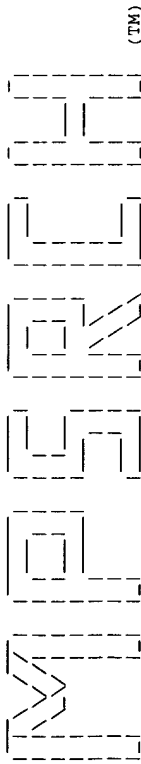
Db 200 SOGLLSC 206

Qy 105 SEGLLAC 111

Search completed: Thu Jul 8 18:18:39 1999

Job time : 18 secs.

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:17:05 1999; MasPar time 3.44 Seconds

Tabular output not generated. 65.831 Million cell updates/sec

Title: >US-09-041-236-2
Description: (105-112) from US09041236.pep (7 of 45)
Perfect Score: 57
Sequence: 1 SEGLIACG 8

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 22.517; Variance 23.231; scale 0.969

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	50	87.7	432	1	SPAL_SALTY PROBABLE ATP SYNTHASE	5.03e-01
2	50	87.7	434	1	FLII_HELPY FLAGELLUM-SPECIFIC ATP	5.03e-01
3	50	87.7	439	1	YSCN_YERPS PROBABLE ATP SYNTHASE	5.03e-01
4	50	87.7	439	1	YSCN_YEREN PROBABLE ATP SYNTHASE	5.03e-01
5	50	87.7	451	1	Y4YI_RHISN PROBABLE ATP SYNTHASE	5.03e-01
6	47	82.5	389	1	YND5_YEAST HYPOTHETICAL 43.8 KD P	2.65e-00
7	46	80.7	279	1	TYSY_LACIA THYMIDYLATE SYNTHASE (4.54e-00
8	46	80.7	302	1	DAPA_PROMA DIHYDRODIPICOLINATE SY	4.54e-00
9	45	78.9	301	1	DAPA_SYNV3 DIHYDRODIPICOLINATE SY	7.70e+00
10	45	78.9	469	1	EXOC_AZOBP PHOSPHOMANNOMUTASE (EC	7.70e+00
11	45	78.9	748	1	Y4Y3_SYNV3 TRNA (GUANINE-N1)-METH	1.29e+01
12	44	77.2	255	1	TRMD_ECOLI NAD/MYCOTHIOI-DEPENDEN	1.29e+01
13	44	77.2	360	1	FADH_ARYME PROBABLE ATP SYNTHASE	1.29e+01
14	44	77.2	430	1	SPAL_SHIFL PROBABLE ATP SYNTHASE	1.29e+01
15	44	77.2	909	1	PSD2_HUMAN 26S PROTEASOME REGULAT	1.29e+01
16	44	77.2	926	1	MEI9_DRONE MEI-9 PROTEIN (MEIOTIC	1.29e+01
17	44	77.2	1131	1	YANC_SCHPO HYPOTHETICAL 128.8 KD	1.29e+01
18	43	75.4	238	1	SFSA_HAEIN SUGAR FERMENTATION STI	2.16e+01
19	43	75.4	240	1	AMDA_ALCBR ARYLALONATE DECARBOXY	2.16e+01
20	43	75.4	245	1	TRMD_HAEIN TRNA (GUANINE-N1)-METH	2.16e+01
21	43	75.4	245	1	TRMD_SALTY TRNA (GUANINE-N1)-METH	2.16e+01
22	43	75.4	255	1	TRMD_SERMA TRNA (GUANINE-N1)-METH	2.16e+01
23	43	75.4	302	1	HYPB_BRAJA HYDROGENASE EXPRESSION	2.16e+01

24	43	75.4	356	1	DHSO_HUMAN SORBITOL DEHYDROGENASE	2.16e+01
25	43	75.4	464	1	P2C2_YEAST PROTEIN PHOSPHATASE 2C	2.16e+01
26	43	75.4	575	1	UL87_EBV PROTEIN B(C)REF1.	2.16e+01
27	43	75.4	607	1	V66K_BWYVF 66.2 KD PROTEIN (ORF 2	2.16e+01
28	43	75.4	634	1	DHSX_YEAST PROBABLE SUCCINATE DEH	2.16e+01
29	43	75.4	640	1	DHSA_YEAST SUCCINATE DEHYDROGENAS	2.16e+01
30	43	75.4	662	1	HEPA_HSV6U DNA HELICASE/PRIMASE C	2.16e+01
31	43	75.4	662	1	HEPA_HSV6Z DNA HELICASE/PRIMASE C	2.16e+01
32	43	75.4	694	1	TRFE_RABIT SEROTRANSFERRIN PRECUR	2.16e+01
33	43	75.4	777	1	L100_ADE41 LATE 100 KD PROTEIN.	2.16e+01
34	43	75.4	1091	1	CIC2_RAT DIHYDROPYRIDINE-SENSIT	2.16e+01
35	43	75.4	1308	1	ERB4_HUMAN ERBB-4 RECEPTOR PROTEI	2.16e+01
36	42	73.7	198	1	HIS5_METJA AMIDOTRANSFERASE HISH	3.55e+01
37	42	73.7	252	1	MODB_AZOVI MOLYBDATE-BINDING PERI	3.55e+01
38	42	73.7	265	1	YG2T_YEAST HYPOTHETICAL 29.7 KD P	3.55e+01
39	42	73.7	267	1	MYB1_HORVU MYB-RELATED PROTEIN HV	3.55e+01
40	42	73.7	280	1	VIE3_HCMVT 30 KD IMMEDIATE-EARLY	3.55e+01
41	42	73.7	365	1	GRA2_BACSU SPORE GERMINATION PROT	3.55e+01
42	42	73.7	410	1	VIE2_HCMVT 45 KD IMMEDIATE-EARLY	3.55e+01
43	42	73.7	522	1	TSAW_RICTS 56 KD TYPE-SPECIFIC AN	3.55e+01
44	42	73.7	770	1	L100_ADE40 LATE 100 KD PROTEIN.	3.55e+01
45	42	73.7	805	1	L100_ADE02 LATE 100 KD PROTEIN.	3.55e+01

ALIGNMENTS

RESULT 1
ID SPAL_SALTY STANDARD; PRT; 432 AA.
AC P39444;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PROBABLE ATP SYNTHASE SPAL/INVC (EC 3.6.1.34).
GN SPAL OR INVC.
OS SALMONELLA TYPHIMURIUM.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SR11 / SL1344;
RX MEDLINE: 94321319.
RA EICHELBERG K., GINOCCHIO C.C., GALAN J.E.;
RT "Molecular and functional characterization of the Salmonella
typhimurium invasion genes invB and invC: homology of invC to the
F0F1 ATPase family of proteins.";
RL J. BACTERIOL. 176:4501-4510(1994).
RN [2]
RP SEQUENCE OF 97-432 FROM N.A.
RX MEDLINE: 94008985.
RA GROISMAN E.A., OCHMAN H.;
RT "Cognate gene clusters govern invasion of host epithelial cells by
Salmonella typhimurium and Shigella flexneri.";
EMBO J. 12:3779-3787(1993).
RN [3]
RP SEQUENCE OF 403-432 FROM N.A.
RC STRAIN=SR11 / SL1344;
RX MEDLINE: 95272391.
RA COLLAZO C., ZIERLER M.K., GALAN J.E.;
RT "Functional analysis of the Salmonella typhimurium invasion genes
invl and invJ and identification of a target of the protein secretion
apparatus encoded in the inv locus.";
MOL. MICROBIOL. 15:25-38(1995).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT ENTRY OF S.TYPHIMURIUM INTO
CULTURED EPITHELIAL CELLS. PROBABLE CATALYTIC SUBUNIT OF A PROTEIN
TRANSLUCASE. MAY ENERGIZE THE PROTEIN EXPORT APPARATUS ENCODED IN
THE INV LOCUS WHICH IS REQUIRED FOR THE SURFACE PRESENTATION OF
DETERMINANTS NEEDED FOR THE ENTRY OF SALMONELLA SPECIES INTO
MAMMALIAN CELLS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: TO OTHER TRANSPORT APPARATUS ATPASES AND TO ATPASE
BETA SUBUNIT.
CC -----
CC This swiss-prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; U08279; G497222; -;
CC EMBL; X73525; G404287; -;
CC EMBL; U10872; G507902; -;
CC PIR; S37304; S37304.
CC STYGENE; SG10464; SPAL.
CC PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
CC PFAM; PF00006; ATP-synt_ab; 1.
CC HYDROLASE; HYDROGEN ION TRANSPORT; ATP SYNTHESIS; ATP-BINDING;
KW VIROLENE.
FT NP_BIND 159 166 ATP (POTENTIAL).
FT MUTAGEN 165 165 K->E: LOSS OF FUNCTION.
FT CONFLICT 291 291 G -> A (IN REF. 2).
FT CONFLICT 388 401 CAKISITIGRCRG -> GENIDNDRAMQMR (IN
FT REF. 2).
FT CONFLICT 407 407 R -> W (IN REF. 2).
SQ SEQUENCE 432 AA; 47253 MW; DC79488F CRC32;

Query Match 87.7%; Score 50; DB 1; Length 432;
Best Local Similarity 71.4%; Pred. No. 5.03e-01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 144 DGLLTGC 150
Qy 106 EGLLAGC 112
:||||:

RESULT 2
ID FLII HELPY STANDARD; PRT; 434 AA.
AC 007025; P94822;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE FLAGELLUM-SPECIFIC ATP SYNTHASE (EC 3.6.1.34).
GN FLII OR HP1420.
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC HELICOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N6;
RX MEDLINE; 97375061.
RA JENKS P.J., FOYNES S., WARD S., WREN B.W.;
RT "A flagellar-specific AtPase (FlII) is necessary for flagellar export
RT in Helicobacter pylori."
RL FEMS MICROBIOL. LETT. 152:205-211(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE; 97394467.
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL NATURE 388:539-547(1997).
CC -!- FUNCTION: PROBABLE CATALYTIC SUBUNIT OF A PROTEIN TRANSLOCASE FOR
CC FLAGELLUM-SPECIFIC EXPORT, OR A PROTON TRANSLOCASE INVOLVED IN
CC LOCAL CIRCUITS AT THE FLAGELLUM.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: TO OTHER TRANSPORT APPARATUS ATPASES AND TO ATPASE

CC BETA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; Y08620; E307987; -;
CC EMBL; AE000642; G2314594; -;
CC TIGR; HP1420;
CC PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
CC PFAM; PF00006; ATP-synt_ab; 1.
CC HSSP; P07677; 1SKY.
CC HYDROLASE; HYDROGEN ION TRANSPORT; ATP SYNTHESIS; ATP-BINDING;
KW TRANSPORT; PROTEIN TRANSPORT; FLAGELLA.
FT NP_BIND 164 171 ATP (POTENTIAL).
FT CONFLICT 15 15 D -> E (IN REF. 1).
FT CONFLICT 115 115 A -> V (IN REF. 1).
FT CONFLICT 140 140 I -> V (IN REF. 1).
FT CONFLICT 323 323 T -> A (IN REF. 1).
FT CONFLICT 367 367 C -> Y (IN REF. 1).
FT CONFLICT 424 424 T -> Q (IN REF. 1).
SQ SEQUENCE 434 AA; 47636 MW; 81EFEE26 CRC32;

Query Match 87.7%; Score 50; DB 1; Length 434;
Best Local Similarity 71.4%; Pred. No. 5.03e-01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 149 DGLLTGC 155
Qy 106 EGLLAGC 112
:||||:

RESULT 3
ID YSCN_YERPS STANDARD; PRT; 439 AA.
AC P40291;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROBABLE ATP SYNTHASE YSCN (EC 3.6.1.34) (YOPS SECRETION ATPASE).
GN YSCN.
OS YERSINIA PSEUDOTUBERCULOSIS.
OC PLASMID PIB1.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC YERSINIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YPIII;
RX MEDLINE; 94222840.
RA BERGMAN T., ERICKSON K., GALYOV E., PERSSON C., WOLF-WATZ H.;
RT "The lcrB (yscN/U) gene cluster of Yersinia pseudotuberculosis is
RT involved in Yop secretion and shows high homology to the spa gene
RT clusters of Shigella flexneri and Salmonella typhimurium."
RL J. BACTERIOL. 176:2619-2626(1994).
CC -!- FUNCTION: IS A COMPONENT OF THE YOP SECRETION MACHINERY USING ATP,
CC AND IT IS EITHER THE ENERGIZER OF THIS MACHINERY, OR PART OF IT.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC BUT CONNECTED TO THE INNER
CC MEMBRANE POSSIBLY THROUGH LCRD (PROBABLE).
CC -!- SIMILARITY: TO OTHER TRANSPORT APPARATUS ATPASES AND TO ATPASE
CC BETA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; U00998; G392946; -;

```
DR EMBL; L23522; G388096; ..
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PFAM; PF00006; ATP-synt_ab; 1.
DR HSP; P09219; ISKY.
KW HYDROLASE; HYDROGEN ION TRANSPORT; ATP SYNTHESIS; ATP-BINDING;
KW TRANSFOSPHATASE; PROTEIN TRANSPORT; VIRULENCE; PLASMID.
FT NP_BIND 169 176 ATP (POTENTIAL).
FT MUTAGEN 169 177 MISSING: ABOLISH ATP-BINDING.
SQ SEQUENCE 439 AA; 47812 MW; 680A7A55 CRC32;

Query Match      87.7%; Score 50; DB 1; Length 439;
Best Local Similarity 71.4%; Pred. No. 5.03e-01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 154 DGLLTCG 160
QY 106 EGLLACG 112

RESULT 4
ID YSCN_YEREN STANDARD; PRT; 439 AA.
AC P40290;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROBABLE ATP SYNTHASE YSCN (EC 3.6.1.34) (YOPS SECRETION ATPASE).
GN YSCN.
OS YERSINIA ENTEROCOLITICA.
OG PLASMID PYVE227.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC YERSINIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KNG22703 / SEROTYPE O:9;
RX MEDLINE; 94179088.
RA WOESTYN S., ALLAUOI A., WATTIAU P., CORNELIS G.R.;
RT "YscN, the putative energizer of the Yersinia Yop secretion
  machinery."
RL J. BACTERIOL. 176:1561-1569(1994).
CC -1- FUNCTION: IS A COMPONENT OF THE YOP SECRETION MACHINERY USING ATP,
  AND IT IS EITHER THE ENERGIZER OF THIS MACHINERY, OR PART OF IT.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC BUT CONNECTED TO THE INNER
  MEMBRANE POSSIBLY THROUGH LCID (PROBABLE).
CC -1- SIMILARITY: TO OTHER TRANSPORT APPARATUS ATPASES AND TO ATPASE
  BETA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U02499; G437202; ..
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PFAM; PF00006; ATP-synt_ab; 1.
DR HSP; P09219; ISKY.
KW HYDROLASE; HYDROGEN ION TRANSPORT; ATP SYNTHESIS; ATP-BINDING;
KW TRANSFOSPHATASE; PROTEIN TRANSPORT; VIRULENCE; PLASMID.
FT NP_BIND 169 176 ATP (POTENTIAL).
FT MUTAGEN 169 177 MISSING: ABOLISH ATP-BINDING.
SQ SEQUENCE 439 AA; 47782 MW; CAE87B72 CRC32;

Query Match      87.7%; Score 50; DB 1; Length 439;
Best Local Similarity 71.4%; Pred. No. 5.03e-01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 154 DGLLTCG 160
QY 106 EGLLACG 112

RESULT 5
ID Y4VI_RHISN STANDARD; PRT; 451 AA.
AC P55717;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROBABLE ATP SYNTHASE Y4VI (EC 3.6.1.34).
GN Y4VI.
OS RHIZOBIUM SP. (STRAIN NGR234).
OG PLASMID SYM PNGR234A.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; RHIZOBIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX FREIBERG C.A., FELLAY R., BAIRIOCH A., BROUGHTON W.J., ROSENTHAL A.,
  PERRET X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL NATURE 387:394-401(1997).
CC -1- FUNCTION: COULD BE INVOLVED IN THE SECRETION OF AN UNKNOWN FACTOR.
CC -1- PROBABLE CATALYTIC SUBUNIT OF THE TRANSLOCASE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: TO OTHER TRANSPORT APPARATUS ATPASES AND TO ATPASE
  BETA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF000107; G2182734; ..
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PFAM; PF00006; ATP-synt_ab; 1.
DR HSP; P09219; ISKY.
KW HYPOTHETICAL PROTEIN; PLASMID; HYDROLASE; HYDROGEN ION TRANSPORT;
  ATP SYNTHESIS; ATP-BINDING; TRANSPORT.
FT NP_BIND 181 188 ATP (POTENTIAL).
SQ SEQUENCE 451 AA; 48765 MW; 21F32DAF CRC32;

Query Match      87.7%; Score 50; DB 1; Length 451;
Best Local Similarity 71.4%; Pred. No. 5.03e-01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 166 DGLLTCG 172
QY 106 EGLLACG 112

RESULT 6
ID YND5_YEAST STANDARD; PRT; 389 AA.
AC P53962;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 43.8 KD PROTEIN IN NCE3-HHT2 INTERGENIC REGION.
GN YND035C OR N2730.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA DUESTERHOEFF A., FLOETH M., FRITZ C., HEUSS-NEITZEL D.,
  HILBERT H., MOESTL D.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; 271311; E239670; -
CC PFAM; PF00400; G-beta; 1.
KW HYPOTHETICAL PROTEIN; ATP-BINDING.
FT NP_BIND 254 261 ATP (POTENTIAL).
SQ SEQUENCE 389 AA; 43848 MW; 60474E7D CRC32;

Query Match 82.5%; Score 47; DB 1; Length 389;
Best Local Similarity 100.0%; Pred. No. 2.65e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 113 GLLACG 118
|||||
QY 107 GLLACG 112

RESULT 7 STANDARD; PRT; 279 AA.
AC P19368;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE THYMIDYLATE SYNTHASE (EC 2.1.1.45) (TS).
GN THYA.
OS LACTOCOCCUS LACTIS (SUBSP. LACTIS) (STREPTOCOCCUS LACTIS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
CC LACTOCOCCUS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-712;
RX MEDLINE; 90358505.
RA ROSS P., O'GARA F., CONDON S.;
RT "Cloning and characterization of the thymidylate synthase gene from
RT Lactococcus lactis subsp. lactis.",
RL APPL. ENVIRON. MICROBIOL. 56:2156-2163(1990).
CC
CC -!- FUNCTION: PROVIDES THE SOLE DE NOVO SOURCE OF DTMP FOR DANA
CC -!- BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: 5,10-METHYLENETETRAHYDROFOLATE + DUMP =>
CC DIHYDROFOLATE + DTMP.
CC -!- PATHWAY: DEOXYRIBONUCLEOTIDE BIOSYNTHESIS.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE SYNTHASE FAMILY. VERY
CC DIVERGENT THYMIDYLATE SYNTHASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M33770; G149513; -
CC PIR; A43797; A43797.
DR PROSITE; PS00091; THYMIDYLATE SYNTHASE; 1.
DR PFAM; PF00303; thymidylat_synt; 1.
DR HSSP; P00469; lTDA.
KW TRANSFERASE; METHYLTRANSFERASE; NUCLEOTIDE BIOSYNTHESIS.
FT ACT_SITE 153 153 BY SIMILARITY.
SQ SEQUENCE 279 AA; 32577 MW; 63C836F3 CRC32;

Query Match 80.7%; Score 46; DB 1; Length 279;
Best Local Similarity 62.5%; Pred. No. 4.54e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 147 TEGLLPCA 154
|||||
QY 105 SEGLLACG 112

RESULT 8 STANDARD; PRT; 302 AA.
ID DAPA_PROMA
AC P49423;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE DIHYDRODIPICOLINATE SYNTHASE (EC 4.2.1.52) (DHDPS).
GN DAPA.
OS PROCHLOROCOCCUS MARINUS.
OC BACTERIA; CYANOBACTERIA; PROCHLOROPHYTES; PROCHLOROCOCCACEAE;
CC PROCHLOROCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COMP 1375;
RA LORENZ M., BOERNER T., HESS W.R.;
RT "Molecular cloning and characterization of a dihydrodipicolinate
RT synthase (DHDPS) gene from the photoautotrophic prokaryote
RT Prochlorococcus marinus CCMP 1375 (Prochlorophyta).";
RL ENDOCYT. CELL RES. 11:59-68(1995).
CC -!- CATALYTIC ACTIVITY: L-ASPARTATE 4-SEMIALDEHYDE + PYRUVATE =>
CC DIHYDRODIPICOLINATE + 2 H(2)O.
CC -!- ENZYME REGULATION: SENSITIVE TO LYSINE INHIBITION
CC (BY SIMILARITY)
CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF DIAMINOPIMELATE AND
CC LYSINE FROM ASPARTATE SEMIALDEHYDE.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DHDPS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; Z68126; E211621; -
DR PROSITE; PS00665; DHDPS_1; 1.
DR PROSITE; PS00666; DHDPS_2; 1.
DR PFAM; PF00701; DHDPS; 1.
DR HSSP; P05640; 1DHP
KW LIASE; DIAMINOPIMELATE BIOSYNTHESIS; LYSINE BIOSYNTHESIS;
KW FEEDBACK-INHIBITION.
FT ACT_SITE 172 172 BY SIMILARITY.
SQ SEQUENCE 302 AA; 31769 MW; 742ADC81 CRC32;

Query Match 80.7%; Score 46; DB 1; Length 302;
Best Local Similarity 62.5%; Pred. No. 4.54e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 46 SDGLIVCG 53
|:|:|:|
QY 105 SEGLLACG 112

RESULT 9 STANDARD; PRT; 301 AA.
ID DAPA_SYNY3
AC Q5513;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DIHYDRODIPICOLINATE SYNTHASE (EC 4.2.1.52) (DHDPS).
GN DAPA OR SLR0650.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 96127529.
RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA SUGIURA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";

RL DNA RES. 2:153-166(1995).
CC -!- CATALYTIC ACTIVITY: L-ASPARTATE 4-SEMIALDEHYDE + PYROUVATE -
CC DIHYDRODIPICOLINATE + 2 H(2)O.
CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF DIAMINOPIMELATE AND
CC LYSINE FROM ASPARTATE SEMIALDEHYDE.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DHDPS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64006; G1001380; -
DR PROSITE; PS00665; DHDPS_1; 1.
DR PROSITE; PS00666; DHDPS_2; 1.
DR PFAM; PF00701; DHDPS; 1.
DR HSSP; P05640; 1DHP
KW LYASE; DIAMINOPIMELATE BIOSYNTHESIS; LYSINE BIOSYNTHESIS.
FT ACT_SITE 170 170 BY SIMILARITY.
SQ SEQUENCE 301 AA; 31821 MW; 6235E80A CRC32;

Query Match 78.9%; Score 45; DB 1; Length 301;
Best Local Similarity 62.5%; Pred. No. 7.70e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 SDGLVVCV 51
I:|:|:|
QY 105 SEGLLACG 112

RESULT 10
ID EXOC_AZOBR STANDARD; PRT; 469 AA.
AC P45632;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PHOSPHOMANNOMUTASE (EC 5.4.2.8) (PMW).
GN EXOC.
OS AZOSPIRILLUM BRASILENSE.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOSPIRILLACEAE;
OC AZOSPIRILLUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SP7;
RA PETERSEN D. J., VANDERLEYDEN J.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: D-MANNOSE 1-PHOSPHATE -> D-MANNOSE 6-PHOSPHATE.
CC -!- SIMILARITY: TO OTHER PHOSPHOGLUCOMUTASES AND PHOSPHOMANNOMUTASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U20583; G695163; -
DR PROSITE; PS00710; PGM_PMW; 1.
DR PFAM; PF00408; PGM_PMW; 1.
KW ISOMERASE; PHOSPHORYLATION.
FT ACT_SITE 106 106 FORMS THE PHOSPHOSERINE INTERMEDIATE
FT (BY SIMILARITY).
SQ SEQUENCE 469 AA; 51130 MW; 45B996B6 CRC32;

Query Match 78.9%; Score 45; DB 1; Length 469;
Best Local Similarity 71.4%; Pred. No. 7.70e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 68 DGLVACG 74
I:|:|:|
QY 106 EGLIACG 112

RESULT 11
ID Y473_SYNY3 STANDARD; PRT; 748 AA.
AC O55168;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 84.2 KD PROTEIN SLR0473.
GN SLR0473.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96127529.
RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA SUGIURA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA RES. 2:153-166(1995).
RN [2]
RP FUNCTION.
RX MEDLINE; 97374241.
RA YEH K.C., LAGARIAS J.C., TEPPERMAN J., QUAIL P.H.;
RT "The phytochromes: a biochemical mechanism of signaling in sight?";
RL UNPUBLISHED RESULTS, CITED BY:
RL QUAIL P.H.;
RL BIOESSAYS 19:571-579(1997).
CC -!- FUNCTION: SEEMS TO FUNCTION AS INFORMATIONAL PHOTORECEPTOR.
CC -!- SIMILARITY: IN THE N-TERMINAL, TO PLANTS PHYTOCHROMES.
CC -!- SIMILARITY: IN THE C-TERMINAL, TO OTHER PROKARYOTIC SENSORY
CC TRANSDUCTION HISTIDINE KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64001; G1001165; -
DR PROSITE; PS00245; PHYTOCHROME_1; FALSE_NEG.
DR PROSITE; PS00046; PHYTOCHROME_2; 1.
DR PFAM; PF00360; phytochrome; 2.
DR PFAM; PF00512; signal; 1.
KW HYPOTHETICAL PROTEIN; SENSORY TRANSDUCTION; TRANSFERASE; KINASE;
KW PHOSPHORYLATION; TRANSCRIPTION REGULATION; PHOTORECEPTOR; PHYTOCHROME;
KW CHROMOPHORE.
FT DOMAIN 20 510 CHROMOPHORE BINDING DOMAIN.
FT DOMAIN 511 748 SENSORY KINASE TRANSMITTER DOMAIN.
FT BINDING 260 260 CHROMOPHORE (BY SIMILARITY).
FT MOD_RES 538 538 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 748 AA; 84232 MW; 5AB71749 CRC32;

Query Match 78.9%; Score 45; DB 1; Length 748;
Best Local Similarity 71.4%; Pred. No. 7.70e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 114 SDGLLVC 120
I:|:|:|
QY 105 SEGLLAC 111

RESULT 12
ID TRMD_ECOLI STANDARD; PRT; 255 AA.
AC P07020;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TRNA (GUANINE-N1)-METHYLTRANSFERASE (EC 2.1.1.31) (M1G-
DE METHYLTRANSFERASE) (TRNA [GM37] METHYLTRANSFERASE).
GN TRMD.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84057772.
RA BYSTROM A.S., HJALMARSSON K.J., WIKSTROM P.M., BJOERK G.R.;
RT "The nucleotide sequence of an Escherichia coli operon containing
RT genes for the tRNA(m1G)methyltransferase, the ribosomal proteins s16
RT and L19 and a 21-k polypeptide.";
RL EMBO J. 2:899-905(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL SCIENCE 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,
RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K., KASAI H.,
RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,
RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,
RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,
RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP CHARACTERIZATION, AND SEQUENCE OF 1-10.
RX MEDLINE: 83108856.
RA HJALMARSSON K.J., BYSTROM A.S., BJORK G.R.;
RT "Purification and characterization of transfer RNA (guanine-
RT 1)methyltransferase from Escherichia coli.";
RL J. BIOL. CHEM. 258:1343-1351(1983).
CC -1- FUNCTION: SPECIFICALLY METHYLATES GUANOSINE-37 IN VARIOUS TRNAs.
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA = S-ADENOSYL-
CC L-HOMOCYSTEINE + TRNA CONTAINING N(1)-METHYLGUANINE.
CC -1- THE SPECIFIC ACTIVITY OF THIS ENZYME INCREASES ONLY SLIGHTLY
CC WITH INCREASED GROWTH RATE.
CC -1- THIS ENZYME IS PRESENT AT CA. 80 MOLECULES/GENOME.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRMD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X01818; G43144; -
DR EMBL: AE000346; G1788959; -
DR EMBL: D90888; G1800012; -
DR PIR: Q00359; XVECG1.
DR ECOGENE: EG11023; TRMD.
KW TRANSFERASE; METHYLTRANSFERASE; TRNA PROCESSING.
SQ SEQUENCE 255 AA; 28422 MW; CAF3D466 CRC32;

Query Match 77.2%; Score 44; DB 1; Length 255;
Best Local Similarity 71.4%; Pred. No. 1.29e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 172 AEGLLDC 178

QY 105 SEGILAC 111
:||||| I
RESULT 13
ID FADH-AMYME STANDARD; PRT; 360 AA.
AC P80094;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NAD/MYCOTHOL-DEPENDENT FORMALDEHYDE DEHYDROGENASE (EC 1.2.1.1-) (MD-
DE FALDH).
OS AMYCOLATOPSIS METHANOLICA.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIADAE;
OC ACTINOMYCETALES; PSEUDONOCARDINEAE; PSEUDONOCARDIACEAE;
OC AMYCOLATOPSIS.
RN [1]
RP SEQUENCE.
RX MEDLINE: 98004265.
RA NORIN A., VAN OPHEM P.W., PERSMA S.R., PERSSON B., DUINE J.A.,
RA JOERNVALL H.;
RT "Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic
RT medium-chain dehydrogenase/reductase, phylogenetically links
RT different eukaryotic alcohol dehydrogenases -- primary structure,
RT conformational modelling and functional correlations.";
RL EUR. J. BIOCHEM. 248:282-289(1997).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-31.
RX MEDLINE: 92283278.
RA VAN OPHEM P.W., VAN BEUENEN J., DUINE J.A.;
RT "NAD-linked, factor-dependent formaldehyde dehydrogenase or trimelic,
RT zinc-containing, long-chain alcohol dehydrogenase from Amycolatopsis
RT methanolica.";
RL EUR. J. BIOCHEM. 206:511-518(1992).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE: 97345657.
RA MISSET-SMITS M., VAN OPHEM P.W., SAKUDA S., DUINE J.A.;
RT "Mycothiol, 1-O-(2'-[N-acetyl-L-cysteinyl]amido-2'-deoxy-alpha-
RT D-glucopyranosyl)-D-xyo-inositol, is the factor of NAD/factor-
RT dependent formaldehyde dehydrogenase.";
RL FEBS LETT. 409:221-222(1997).
CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + MYCOTHOL + NAD(+) -
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- COFACTOR: EACH SUBUNIT BINDS 6 ATOMS OF ZINC.
DR PIR: S23181; S23181.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW OXIDOREDUCTASE; NAD; ZINC.
FT METAL 41 41 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 62 62 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 89 89 SUBSTRATE-BINDING (PROBABLE).
FT METAL 92 92 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 95 95 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 98 98 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 106 106 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 161 161 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 360 AA; 37757 MW; 92D9B367 CRC32;

Query Match 77.2%; Score 44; DB 1; Length 360;
Best Local Similarity 83.3%; Pred. No. 1.29e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 157 GLLGCG 162
:||||| I
QY 107 GLLACG 112

RESULT 14
ID SPAL-SHIFL STANDARD; PRT; 430 AA.
AC P35531;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)

15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PROBABLE ATP SYNTHASE SPAL/MXIB (EC 3.6.1.34).
GN SPAL OR SPA47 OR MXIB.
OS SHIGELLA FLEXNERI, AND SHIGELLA SONNEI.
OG PLASMID 210 KB INVASION PWR100, AND PLASMID 230 KB PMYSH6000.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SHIGELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-S.FLEXNERI; STRAIN=M90T; PLASMID=PWR100;
RX MEDLINE: 92193289.
RA VENKATESAN M.M., BUYSSE J.M., OAKS E.V.;
RT "Surface presentation of Shigella flexneri invasion plasmid antigens
RT requires the products of the spa locus.";
RL J. BACTERIOL. 174:1990-2001(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-S.FLEXNERI; STRAIN=2A; PLASMID=PMYSH6000;
RX MEDLINE: 93224456.
RA SASAKAWA C., KOMATSU K., TOBE T., SUZUKI T., YOSHIKAWA M.;
RT "Eight genes in region 5 that form an operon are essential for
RT invasion of epithelial cells by Shigella flexneri 2a.";
RL J. BACTERIOL. 175:2334-2346(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-S.SONNEI; STRAIN=HW383;
RA ARAKAWA E., KATO J.I., ITO K.I., WATANABE H.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: REQUIRED FOR SURFACE PRESENTATION OF INVASION PLASMID
CC ANTIGENS. PROBABLE CATALYTIC SUBUNIT OF A PROTEIN TRANSLOCASE.
CC REQUIRED FOR INVASION AND FOR SECRETION OF THE THREE IPA PROTEINS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: TO OTHER TRANSPORT APPARATUS ATPASES AND TO ATPASE
CC BETA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M81458; G152792; -.
CC DR EMBL: D13663; D1003330; -.
CC DR EMBL: D50601; G829076; -.
CC DR PIR: C42284; C42284.
CC DR PIR: S27680; S27680.
CC DR PIR: C49846; C49846.
CC DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
CC DR PFAM: PF00006; ATP-synt_ab; 1.
CC DR HSSP: P09219; 1SKY.
CC DR PLASMID: VIRULENCE; HYDROLASE; HYDROGEN ION TRANSPORT; ATP SYNTHESIS;
KW ATP-BINDING; TRANSPORT; PROTEIN TRANSPORT.
FT NP_BIND 159 165 ATP (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47518 MW; 813DE1C2 CRC32;

Query Match 77.2%; Score 44; DB 1; Length 430;
Best Local Similarity 57.1%; Pred. No. 1.29e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 144 DSSLTCG 150
QY 106 EGLLACG 112
:|||||
RESULT 15
ID PSD2_HUMAN STANDARD; PRT; 909 AA.
AC Q13200; Q12932; Q15321;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 26S PROTEASOME REGULATORY SUBUNIT S2 (P97) (TUMOR NECROSIS FACTOR TYPE

1 RECEPTOR ASSOCIATED PROTEIN 2) (55.11 PROTEIN).
DE PSMD2 OR TRAP2.
GN HOMO SAPIENS (HUMAN).
OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FIBROSARCOMA;
RX MEDLINE: 96370847.
RA TSURUMI C., SHIMIZU Y., SAEKI M., KATO S., DEMARTINO G.N.,
RA SLAUGHTER C.A., FUJIMURO M., YOKOSAWA H., YAMASAKI M., HENDIL K.B.,
RA TOH-E A., TANAHASHI N., TANAKA K.;
RT "cDNA cloning and functional analysis of the p97 subunit of the 26S
RT proteasome, a polypeptide identical to the type-1
RT tumor-necrosis-factor-receptor-associated protein-2/55.11.";
RL EUR. J. BIOCHEM. 239:912-921(1996).
RN [2]
RP SEQUENCE OF 58-909 FROM N.A.
RX MEDLINE: 95181307.
RA SONG H.Y., DUNBAR J.D., ZHANG Y.X., GUO D., DONNER D.B.;
RT "Identification of a protein with homology to hsp90 that binds the
RT type 1 tumor necrosis factor receptor.";
RL J. BIOL. CHEM. 270:3574-3581(1995).
RN [3]
RP SEQUENCE OF 10-909 FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 95324729.
RA BOLDIN M.P., METT I.L., WALLACH D.;
RT "A protein related to a proteasomal subunit binds to the
RT intracellular domain of the p55 TNF receptor upstream to its 'death
RT domain'.";
RL FEBS LETT. 367:39-44(1995).
CC -!- FUNCTION: ACTS AS A REGULATORY SUBUNIT OF THE 26 PROTEASOME. WHICH
CC IS INVOLVED IN THE ATP-DEPENDENT DEGRADATION OF UBIQUITINATED
CC PROTEINS.
CC -!- FUNCTION: BINDS TO THE INTRACELLULAR DOMAIN OF TUMOR NECROSIS
CC FACTOR TYPE 1 RECEPTOR.
CC -!- SIMILARITY: BELONGS TO THE PROTEASOME SUBUNIT S2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D78151; G1060888; -.
CC DR EMBL: U12596; G687239; -.
CC DR EMBL: X86446; G1008089; -.
CC KW PROTEASOME; TRANSMEMBRANE.
FT TRANSMEM 514 534 POTENTIAL.
FT CONFLICT 10 10 P -> R (IN REF. 3).
FT CONFLICT 60 60 V -> A (IN REF. 1).
FT CONFLICT 226 226 Y -> S (IN REF. 2).
FT CONFLICT 260 260 S -> T (IN REF. 3).
FT CONFLICT 281 283 IFT -> SS (IN REF. 2).
FT CONFLICT 415 415 G -> A (IN REF. 1).
FT CONFLICT 731 731 M -> MGM (IN REF. 2).
FT CONFLICT 900 909 FGRTPIWISK -> LRKNENYDL (IN REF. 1).
SQ SEQUENCE 909 AA; 100216 MW; DFFFA790 CRC32;

Query Match 77.2%; Score 44; DB 1; Length 909;
Best Local Similarity 75.0%; Pred. No. 1.29e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 442 SGALLACG 449
QY 105 SEGLLACG 112
:|||||
Search completed: Thu Jul 8 18:17:12 1999
Job time : 7 secs.

MAISEH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:17:33 1999; MasPar time 5.03 Seconds
Tabular output not generated. 86.815 Million cell updates/sec

Title: >US-09-041-236-2
Description: (105-112) from US09041236.ppt (7 of 45)
Perfect Score: 57
Sequence: 1 SEGLLACG 8

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 21.694; Variance 25.219; scale 0.860

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	57	100.0	666	4	075326 SEMAPHORIN L.	4.20e-02
2	54	94.7	653	14	064906 SIMILAR TO GENBANK ACC	2.20e-01
3	52	91.2	196	4	075594 PEPTIDOGLYCAN RECOGNIT	6.44e-01
4	51	89.5	4123	4	075851 WUGSC:HLDJ0751H13.1 PR	1.09e+00
5	50	87.7	290	2	P71234 SEPB (FRAGMENT).	1.84e+00
6	50	87.7	434	2	054374 ATPASE.	1.84e+00
7	50	87.7	439	2	068595 YOPS SECRETION ATPASE	1.84e+00
8	50	87.7	440	2	030538 PSCN.	1.84e+00
9	50	87.7	444	2	068539 BSCN.	1.84e+00
10	50	87.7	446	2	085632 L0034.	1.84e+00
11	50	87.7	446	2	052140 ESCN.	1.84e+00
12	50	87.7	450	2	P72272 PUTATIVE ATPASE.	1.84e+00
13	50	87.7	465	2	030438 YUP8H12R.21. PROTEIN.	8.49e+00
14	47	82.5	327	10	064534 GLUCOSYL TRANSFERASE.	8.49e+00
15	47	82.5	457	10	P93709 ATP/GTP BINDING PROTEI	8.49e+00
16	47	82.5	587	2	033229 KIAA0172 PROTEIN (FRAG	8.49e+00
17	47	82.5	1307	4	Q14678 HYPOTHETICAL 7.8 KD PR	1.39e+01
18	46	80.7	73	4	000250 F54B8.1 PROTEIN.	1.39e+01
19	46	80.7	150	5	045558 FROM BASES 2434445 TO	1.39e+01
20	46	80.7	392	2	P77549	

21	46	80.7	406	2	035033 YLOI PROTEIN.	1.39e+01
22	45	78.9	88	11	062821 AXONEMAL DYNEIN HEAVY	2.27e+01
23	45	78.9	161	11	063170 DYNEIN-LIKE PROTEIN 7	2.27e+01
24	45	78.9	164	4	Q15372 HYPOTHETICAL 18.5 KD P	2.27e+01
25	45	78.9	169	4	000433 AXONEMAL DYNEIN HEAVY	2.27e+01
26	45	78.9	204	11	008823 AXONEMAL DYNEIN HEAVY	2.27e+01
27	45	78.9	277	1	058047 277AA LONG HYPOTHETICA	2.27e+01
28	45	78.9	319	2	P72312 NITRILASE REGULATOR.	2.27e+01
29	45	78.9	368	2	088007 PUTATIVE GLYCOSYL TRAN	2.27e+01
30	45	78.9	393	11	088371 SEMAPHORIN L (FRAGMENT	2.27e+01
31	45	78.9	1331	10	049549 HYPOTHETICAL 148.1 KD	2.27e+01
32	45	78.9	2861	4	Q13458 TRIO.	2.27e+01
33	45	78.9	3038	4	Q75962 SIMILAR TO PBCV-1 CVU1	3.67e+01
34	44	77.2	180	14	Q84444 DNA-BINDING PROTEIN.	3.67e+01
35	44	77.2	236	10	Q38851 KOZEV.5 PROTEIN.	3.67e+01
36	44	77.2	278	5	017254 348AA LONG HYPOTHETICA	3.67e+01
37	44	77.2	348	1	058161 PUTATIVE DEHYDROGENASE	3.67e+01
38	44	77.2	361	2	053533 ENVELOPE POLYPROTEIN G	3.67e+01
39	44	77.2	425	14	Q06513 PUTATIVE 3-BETA HYDROX	3.67e+01
40	44	77.2	437	10	Q22856 ENV. PROTEIN (FRAGMENT)	3.67e+01
41	44	77.2	461	14	Q77990 APICAL MEMBRANE ANTIGE	3.67e+01
42	44	77.2	557	5	Q26225 P67.	3.67e+01
43	44	77.2	591	4	Q13041 GP160.	3.67e+01
44	44	77.2	750	14	Q74105 MOLD1.2 PROTEIN.	3.67e+01
45	44	77.2	769	5	Q17224	

ALIGNMENTS

RESULT 1						
ID 075326	PRELIMINARY;	PRT;	666	AA.		
AC 075326:						
DT 01-NOV-1998	(TREMBLREL. 08, CREATED)					
DT 01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)					
DT 01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)					
DE SEMAPHORIN L.						
GN SEMAL.						
OS HOMO SAPIENS (HUMAN).						
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;						
CC CATARRHINI; HOMINIDAE; HOMO.						
RN [1]						
RP SEQUENCE FROM N.A.						
RX MEDLINE: 98389619.						
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;						
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA						
RL viruses.";						
RT GENOMICS 51:340-350(1998).						
DR EMBL: AF030698; G3523115; "						
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;						

Query Match 100.0%; Score 57; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 4.20e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 137 SEGLLACG 144

Qy 105 SEGLLACG 112

RESULT 2
ID 064906
AC 064906;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALCELAEPHINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
CC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RX MEDLINE: 97201573.

RA ENSER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene."
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RX MEDLINE: 97404659.
RA ENSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome."
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RA ENSER A., PFLANZ R., FLECKENSTEIN B.;
RT SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U18243; G1000717; -.
DR EMBL: AF005370; G2337970; -.
SQ SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match 94.7%; Score 54; DB 14; Length 653;
Best Local Similarity 75.0%; Pred. No. 2.20e-01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 159 TDGLLACG 166
:::|||||
Qy 105 SEGLLACG 112

RESULT 3
ID 075594 PRELIMINARY; PRT; 196 AA.
AC 075594;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PEPTIDOGLYCAN RECOGNITION PROTEIN PRECURSOR.
GN PGRP.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BONE MARROW;
RX MEDLINE: 98374308.
RA KANG D., LIU G., LUNDSTROM A., GELIUS E., STEINER H.;
RT "A peptidoglycan recognition protein in innate immunity conserved
RT from insects to humans."
RL PROC. NATL. ACAD. SCI. U.S.A. 95:10078-10082(1998).
DR EMBL: AF076483; G3342533; -.
KW SIGNAL.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 196 PEPTIDOGLYCAN RECOGNITION PROTEIN.
SQ SEQUENCE 196 AA; 21731 MW; DC864D63 CRC32;

Query Match 91.2%; Score 52; DB 4; Length 196;
Best Local Similarity 75.0%; Pred. No. 6.44e-01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 148 AOGLLACG 155
:::|||||
Qy 105 SEGLLACG 112

RESULT 4
ID 075851 PRELIMINARY; PRT; 4123 AA.
AC 075851;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE WUGSC: H_DJ0751H13.1 PROTEIN (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]

RP SEQUENCE FROM N.A.
RA LEONARD S., GRAVES T., STROMMATT C.;
RT "The sequence of Homo sapiens PAC clone DJ0751H13."
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.H.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AC004877; G3638957; -.
DR PROSITE; PS01209; LDLRA_1; 9.
KW GLYCOPROTEIN.
FT NON_TER 1 1
SQ SEQUENCE 4123 AA; 434981 MW; OD93C3B8 CRC32;

Query Match 89.5%; Score 51; DB 4; Length 4123;
Best Local Similarity 85.7%; Pred. No. 1.09e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 2467 EGLLSCG 2473
||||:|
Qy 106 EGLLACG 112

RESULT 5
ID P71234 PRELIMINARY; PRT; 290 AA.
AC P71234;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEPB (FRAGMENT).
GN SEPB.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E2348 /69 (ENTEROPATHOGENIC E. COLI);
RX MEDLINE: 95372406.
RA JARVIS K.G., GIRON J.A., JERSE A.E., MCDANIEL T.K., DONNENBERG M.S.,
RA KAPER J.B.;
RT "Enteropathogenic Escherichia coli contains a putative type III
RT secretion system necessary for the export of proteins involved in
RT attaching and effacing lesion formation."
RL PROC. NATL. ACAD. SCI. U.S.A. 92:7996-8000(1995).
DR EMBL: Z49933; E264337; -.
DR PFAM; PF00006; ATP-synt_ab; 1.
FT NON_TER 290 290
SQ SEQUENCE 290 AA; 31555 MW; CF1EE1D9 CRC32;

Query Match 87.7%; Score 50; DB 2; Length 290;
Best Local Similarity 71.4%; Pred. No. 1.84e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 162 DGLLTCG 168
||||:|
Qy 106 EGLLACG 112

RESULT 6
ID O54374 PRELIMINARY; PRT; 434 AA.
AC O54374;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE ATPASE.
GN FLII.
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC HELICOBACTER.

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CCUG 17874;
RA PORWOLLIK S., O'TOOLE P.W.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U7584; G2731646; -.
SQ SEQUENCE 434 AA; 47592 MW; 221EA346 CRC32;

Query Match      87.7%; Score 50; DB 2; Length 434;
Best Local Similarity 71.4%; Pred. No. 1.84e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 149 DGLLTCG 155
      :|||:|
Qy 106 EGLLAGC 112

RESULT 7
ID O68695 PRELIMINARY; PRT; 439 AA.
AC O68695;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE YOPS SECRETION ATPASE YSCN HOMOLOG.
GN YSCN.
OS YERSINIA PESTIS.
OG PLASMID PCDL.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC YERSINIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KIM;
RA HU P., ELLIOTT J., MCCREADY P., SKOWRONSKI E., GARNES J.,
RA KOBAYASHI A., CARRANO A.V., BRUBAKER R., GARCIA E.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF053946; G2996241; -.
KW PLASMID.
SQ SEQUENCE 439 AA; 47800 MW; 3B6BF752 CRC32;

Query Match      87.7%; Score 50; DB 2; Length 439;
Best Local Similarity 71.4%; Pred. No. 1.84e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 154 DGLLTCG 160
      :|||:|
Qy 106 EGLLAGC 112

RESULT 8
ID O30538 PRELIMINARY; PRT; 440 AA.
AC O30538;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE PSN.
GN PSN.
OS PSEUDOMONAS AERUGINOSA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC PSEUDOMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-386;
RA ROONEY P.J., YAHN T.L., FRANK D.W.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF010151; G2459981; -.
DR PFAM: PF00006; ATP-synt.ab; 1.
SQ SEQUENCE 440 AA; 47934 MW; 8B1D64DF CRC32;

Query Match      87.7%; Score 50; DB 2; Length 440;
Best Local Similarity 71.4%; Pred. No. 1.84e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 155 DGLLTCG 161

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CCUG 112
      :|||:|
Qy 106 EGLLAGC 112

RESULT 9
ID O68539 PRELIMINARY; PRT; 444 AA.
AC O68539;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE BSCN.
GN BSCN.
OS BORDETTELLA BRONCHISEPTICA.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; ALCALIGENACEAE;
OC BORDETTELLA.
RN [1]
RP SEQUENCE FROM N.A.
RA YUK M.H., HARVILL E.T., MILLER J.F.;
RL MOL. MICROBIOL. 0:0-0(1998).
DR EMBL: AF049488; G2935537; -.
SQ SEQUENCE 444 AA; 48310 MW; 37F615D7 CRC32;

Query Match      87.7%; Score 50; DB 2; Length 444;
Best Local Similarity 71.4%; Pred. No. 1.84e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 156 DGLLTCG 162
      :|||:|
Qy 106 EGLLAGC 112

RESULT 10
ID O85632 PRELIMINARY; PRT; 446 AA.
AC O85632;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE L0034.
GN ESCN.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EDL933;
RX MEDLINE; 98339885.
RA PERNA N.T., MAYHEW G.F., POSFAI G., ELLIOTT S., DONNENBERG M.S.,
RA KAPER J.B., BLATTNER F.R.;
RT "Molecular evolution of a pathogenicity island from enterohemorrhagic
RL escherichia coli O157:H7.";
RL INFECT. IMMUN. 66:3810-3817(1998).
DR EMBL: AF071034; G3414902; -.
SQ SEQUENCE 446 AA; 48816 MW; AB96B8B4 CRC32;

Query Match      87.7%; Score 50; DB 2; Length 446;
Best Local Similarity 71.4%; Pred. No. 1.84e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 162 DGLLTCG 168
      :|||:|
Qy 106 EGLLAGC 112

RESULT 11
ID O52140 PRELIMINARY; PRT; 446 AA.
AC O52140;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE ESCN.
GN ESCN.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
```

```
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RA ELIOTT S.J., WAINWRIGHT L.A., MCDANIEL T.K., JARVIS K.G., DENG Y.K.,
RA LAI L.C., MCNAMARA B.P., DONNENBERG M.S., KAPER J.B.,
RA MOL. MICROBIOL. 0:0-0(1998).
DR EMBL; AF022236; G2865290; -.
SQ SEQUENCE 446 AA; 48858 MW; A8E4DAC0 CRC32;

Query Match      87.7%; Score 50; DB 2; Length 446;
Best Local Similarity 71.4%; Pred. No. 1.84e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 162 DGLLTCG 168
   :|||:|
QY 106 EGLLACG 112

RESULT 12
ID P72272 PRELIMINARY; PRT; 450 AA.
AC P72272;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE ATPASE.
GN HRCN.
OS RHIZOBIUM FREDDII.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; SINORHIZOBIUM.
RN [1]
RP SEQUENCE OF 1-116 FROM N.A.
RX MEDLINE; 94018604.
RA MEINHARDT L.W., KRISHNAN H.B., BALATTI P.A., PUEPKE S.G.;
RT "Molecular cloning and characterization of a sym plasmid locus that
RT regulates cultivar-specific nodulation of soybean by Rhizobium fredii
RT USDA257.";
RL MOL. MICROBIOL. 17:923-933(1995).
RN [2]
RP SEQUENCE OF 1-116 FROM N.A.
RX MEDLINE; 96123434.
RA KOVACS L.G., BALATTI P.A., KRISHNAN H.B., PUEPKE S.G.;
RT "Transcriptional organization and expression of no1xwbtuv, a locus
RT that regulates cultivar-specific nodulation of soybean by Rhizobium
RT fredii USDA257.";
RN [3]
RP SEQUENCE FROM N.A.
RA KRISHNAN H.B., PUEPKE S.G.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RA HE X.T., KRISHNAN H.B., PUEPKE S.G.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; L12251; G1648933; -.
DR PFAM; PF00006; ATP-synt_ab; 1.
SQ SEQUENCE 450 AA; 49535 MW; 5D31A8EB CRC32;

Query Match      87.7%; Score 50; DB 2; Length 450;
Best Local Similarity 71.4%; Pred. No. 1.84e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 165 DGLLTCG 171
   :|||:|
QY 106 EGLLACG 112

RESULT 13
ID O30438 PRELIMINARY; PRT; 465 AA.
AC O30438;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE ATPASE.
```

```
GN YSAL.
OS YERSINIA ENTEROCOLITICA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC YERSINIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8081;
RA CARLSON S., PIERSON D.E.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF005744; G2352532; -.
DR PFAM; PF00006; ATP-synt_ab; 1.
SQ SEQUENCE 465 AA; 50213 MW; 55F93EF8 CRC32;

Query Match      87.7%; Score 50; DB 2; Length 465;
Best Local Similarity 71.4%; Pred. No. 1.84e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 145 DGLLTCG 151
   :|||:|
QY 106 EGLLACG 112

RESULT 14
ID O64534 PRELIMINARY; PRT; 327 AA.
AC O64534;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE YUP8H12R.21 PROTEIN.
GN YUP8H12R.21.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA THEOLOGIS A., VYSOTSKAIA V.S., OSBORNE B.I., SCHWARTZ J.R.,
RA FEDERSPIEL N.A., KWAN A., TORIUMI M., YU G., OJI, O., ARAUJO R.,
RA CHUNG E., DEWER K., DIETRICH F., ECKER J.R., MARZIALI A., OEFNER P.,
RA DAVIS R.W.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AC002986; G3152563; -.
SQ SEQUENCE 327 AA; 37643 MW; BCFD10D3 CRC32;

Query Match      82.5%; Score 47; DB 10; Length 327;
Best Local Similarity 75.0%; Pred. No. 8.49e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 78 NEGLRCG 85
   :|||:|
QY 105 SEGLLACG 112
```



```

RESULT 15
ID P93709 PRELIMINARY; PRT: 467 AA.
AC P93709;
DT 01-MAY-1997 (TREMBREL. 03, CREATED)
DT 01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBREL. 09, LAST ANNOTATION UPDATE)
DE GLUCOSYL TRANSFERASE.
GN JIGT.
OS NICOTIANA TABACUM (COMMON TOBACCO).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; NICOTIANA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BY-2; TISSUE-SUSPENSION CULTURE;
RA KOJIMA H., HASHIZUME K., IMANISHI S., NAKAMURA K.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AB000623; D1019901; -.
DR PROSITE: PS00375; UDPGT; 1.
DR PFAM: PF00201; UDPGT; 1.
DR MENDEL: 9421; NICTa:1525; 3.
KW TRANSFERASE.
SQ SEQUENCE 467 AA: 52989 MW: 3726EB31 CRC32;

Query Match 82.5%; Score 47; DB 10; Length 467;
Best Local Similarity 75.0%; Pred. No. 8.49e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 298 SOGLLKCG 305
Oy 105 SEGLLACG 112

```

Search completed: Thu Jul 8 18:18:01 1999
 Job time : 28 secs.

THIS PAGE BLANK (USPTO)

MAQREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:21:52 1999; MasPar time 10.33 Seconds
Tabular output not generated. 14.417 Million cell updates/sec

Title: >US-09-041-236-2
Description: (116-122) from US09041236.pep (8 of 45)
Perfect Score: 66
Sequence: 1 RHPCSWN 7
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 16.136; Variance 51.948; scale 0.311

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	53	80.3	530	8	R43024	Bx1 gene product.	4.86e+01
2	50	75.8	123	31	W48351	Human breast cancer r	9.92e+01
3	50	75.8	242	32	W58379	Human immunomodulator	9.92e+01
4	48	72.7	28	18	R91384	MEH1 library derived	1.59e+02
5	48	72.7	129	8	R42832	Cytochrome P450 homol	1.59e+02
6	48	72.7	1805	5	R27204	Rat nestin.	1.59e+02
7	48	72.7	1805	11	R60126	Rat nestin protein is	1.59e+02
8	47	71.2	586	23	W10243	Cytomegalovirus UL84	2.00e+02
9	46	69.7	95	3	R12603	SIB 121 intestinal mu	2.52e+02
10	46	69.7	163	21	W11283	Human breast specific	2.52e+02
11	45	68.2	383	1	P91395	Peptide expressed by	3.17e+02
12	45	68.2	384	10	R51884	KAPA synthase encoded	3.17e+02
13	45	68.2	502	2	R10274	Simian immunodefice	3.17e+02
14	45	68.2	1565	25	W18305	Photorehabdus luminesc	3.17e+02
15	45	68.2	1565	31	W56568	Toxin TccB, encoded b	3.17e+02
16	45	68.2	2016	30	W23994	Human hH1 sodium chan	3.17e+02

17	45	68.2	2019	13	R67913	Cardiac sodium channe	3.17e+02
18	45	68.2	2020	2	R05584	Cardiac sodium channe	3.17e+02
19	45	68.2	2115	32	W52276	Rubella virus RA27/3	3.17e+02
20	45	68.2	2205	15	R79048	Infectious rubella vi	3.17e+02
21	45	68.2	3011	16	R95021	Hepatitis GB virus (H	3.17e+02
22	44	66.7	16	23	W24884	Predatory cone snail	3.99e+02
23	44	66.7	16	14	R75277	A-lineage conotoxin C	3.99e+02
24	44	66.7	16	20	W12739	A-lineage conotoxin p	3.99e+02
25	44	66.7	17	2	R12840	HIV protecting vaccin	3.99e+02
26	44	66.7	137	4	R20598	GAG gene p12 protein.	3.99e+02
27	44	66.7	137	2	P70550	HIV virus anti-p12 an	3.99e+02
28	44	66.7	315	4	R22214	Sequence of interleuk	3.99e+02
29	44	66.7	323	9	R47177	Human autocrine motil	3.99e+02
30	44	66.7	395	14	R80288	galk gene of S. livid	3.99e+02
31	44	66.7	395	8	R41530	S. lividans galactose-	3.99e+02
32	44	66.7	398	4	R22212	Sequence of interleuk	3.99e+02
33	44	66.7	415	4	R22217	Sequence of interleuk	3.99e+02
34	44	66.7	415	4	R22211	Sequence of interleuk	3.99e+02
35	44	66.7	506	1	P80802	Sequence encoded by g	3.99e+02
36	44	66.7	510	4	R22365	SIVmac239 gag gene pr	3.99e+02
37	44	66.7	516	1	P80808	Sequence of gag prote	3.99e+02
38	44	66.7	521	23	W13053	HIV-2 provirus-encode	3.99e+02
39	44	66.7	521	1	P80807	Sequence of gag prote	3.99e+02
40	44	66.7	522	1	P70554	HIV virus gag gene.	3.99e+02
41	44	66.7	522	1	P82676	GAGRODN sequence from	3.99e+02
42	44	66.7	522	2	R04024	Gag gene product of c	3.99e+02
43	44	66.7	522	1	P80801	Sequence encoded by o	3.99e+02
44	44	66.7	3077	1	P93283	Sequence of clone HIV	3.99e+02
45	44	66.7	3211	1	P81769	Sequence encoded by t	3.99e+02

ALIGNMENTS

RESULT 1

ID R43024 standard; protein; 530 AA.
AC: R43024;
DT 17-MAY-1994 (first entry)
DE Bx1 gene product.
KW Bx1; resistance; plants; benzoxazine; biosynthesis; alleles;
KW European corn borer; pest; vector; clone.
OS Zea mays.
PN W09322441-A.
PD 11-NOV-1993.
PF 23-APR-1992; E00905.
PR 23-APR-1992; WO-E00905.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Frey M, Gierl A, Peterson PA, Saedler H, Sommer H;
DR WPI: 93-368800/46.
DR N-PSDB: Q50511.
PT DNA sequence of Bx1 gene - used to confer resistance on plants
PT with low or no levels of benzoxazine(s)
PS Claim 1; Fig 2; 28pp; English.
CC The protein is involved in the biosynthesis of benzoxazines,
CC which are used by plants as a poison / deterrent on insects and
CC microorganisms. The protein can be expressed in transformed plants,
CC enhancing their ability to combat infection.
SQ Sequence 530 AA;

Query Match 80.3%; Score 53; DB 8; Length 530;
Best Local Similarity 57.1%; Pred. No. 4.86e+01;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 425 rdptcwd 432
| :| :| :|
Qy 116 RHPCSWN 122

RESULT 2

ID W48351 standard; Protein; 123 AA.
AC W48351;
DT 14-AUG-1998 (first entry)
DE Human breast cancer related protein BCRB2.
KW L-oncogene; diagnosis; treatment; ovarian cancer; antibody; antisense.

OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Binding_site 5..10
 FT FT [note="haem iron binding site"]
 FT Domain 26..46
 FT FT [note="Transmembrane helix"]
 FT Domain 79..108
 FT FT [note="Transmembrane helix"]
 PN WO9807851-A2.
 PN 26-FEB-1998.
 PD 22-AUG-1997; E04600.
 PR 22-AUG-1996; CA-183500.
 PA (BERG/) BERGMANN J E.
 PA (PREDD/) PREDDIE E R.
 PI Bergmann JE, Preddie ER;
 DR WPI: 98-169156/15.
 DR N-PSDB: V17747.
 PT Human breast cancer related genes - used for diagnosis,
 PT pre-symptomatic detection and therapy of breast and ovarian cancers
 PS Claim 2; Fig 3e; 96pp; English.
 CC The proteins encoded by human breast cancer related genes, L-oncogenes
 CC (V17738-V17753) can be used in the diagnosis and treatment of breast and
 CC ovarian cancer. Antibodies and multiple antigenic peptide epitopes can
 CC be used to detect the presence of the proteins. The antibodies can also
 CC be used to block the activity of the protein. Antisense molecules can be
 CC used to prevent expression of the proteins.
 SQ Sequence 123 AA;
 Query Match 75.8%; Score 50; DB 31; Length 123;
 Best Local Similarity 56.7%; Pred. No. 9.92e+01;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 65 hpkcw 70
 QY 117 HPSCWN 122
 RESULT 3
 ID W58379 standard; Protein; 242 AA.
 AC W58379;
 DT 08-SEP-1998 (first entry)
 DE Human immunomodulatory protein BL-1 3.
 KW Human; BL-1; immunotherapy; vaccine; immunomodulatory protein; HIV;
 KW infection; cancer; autoimmune disease.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 1..242
 FT [note="the protein is decoded from the third reading
 frame of BL-1 cDNA and so contains stop codons
 (i.e. all X's are stop codons)"]
 PN WO9817799-A1.
 PN 30-APR-1998.
 PD 23-OCT-1997; U19502.
 PR 23-OCT-1996; US-028613.
 PA (APOL-) APOLLON INC.
 PA (UYPE-) UNIV PENNSYLVANIA.
 PI Bagarazzi ML, Boyer JD, Kim JJ, Wang B, Weiner DB;
 DR WPI: 98-261495/23.
 DR N-PSDB: V30915.
 PT New compositions for immuno-therapy and protection - comprise
 PT nucleotide sequences encoding an immuno-modulating protein and an
 PT antigen, used for e.g. infections, cancer or auto-immune diseases
 PS Example 6; Fig 14; 136pp; English.
 CC The present sequence represents the protein from the third reading
 CC frame of human immunomodulatory protein BL-1 encoding cDNA (given in
 CC V30915). The present invention describes a plasmid which comprises a
 CC nucleotide sequence (NS) that encodes: (a) an immunomodulating protein
 CC selected from interleukin (IL)-12, granulocyte-macrophage colony
 CC stimulating factor (GM-CSF), IL-1, tumour necrosis factor (TNF)-alpha,
 CC TNF-beta, IL-2, IL-4, IL-5, IL-10, IL-15, IL-18 and BL-1 operably
 CC linked to regulatory elements. Products of the present invention can be
 CC used to induce an immune response to an antigen such as a pathogen
 CC antigen, a hyperproliferative disease-associated antigen, and antigen

CC linked to cells associated with autoimmune diseases or an allergen.
 CC They can be used for immunotherapy or to provide a protective immune
 CC response. In particular, they can be used for treating subjects with
 CC an allergic reaction, pathogen infection, hyperproliferative disease
 CC such as cancer or psoriasis or autoimmune diseases e.g. rheumatoid
 CC arthritis, multiple sclerosis, Sjogren's syndrome, sarcoidosis, insulin
 CC dependent diabetes mellitus, autoimmune thyroiditis, ankylosing
 CC spondylitis, scleroderma, polymyositis, dermatomyositis, psoriasis,
 CC vasculitis, Wegener's granulomatosis, Crohn's disease and ulcerative
 CC colitis, Grave's disease, autoimmune thrombocytopenia, asthma and
 CC pernicious anaemia.
 SQ Sequence 242 AA;
 Query Match 75.8%; Score 50; DB 32; Length 242;
 Best Local Similarity 71.4%; Pred. No. 9.92e+01;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 9 rppvcmw 15
 QY 116 RHPSCWN 122

RESULT 4
 ID R91384 standard; Peptide; 28 AA.
 AC R91384;
 DT 10-OCT-1996 (first entry)
 DE ME#1 library derived peptide, 950406-11.
 KW C1s-element; ligand; highly specific DNA binding domain; HSDB; human;
 KW H2-kappa-B; regulatory domain; murine; homologue; IL-6; interleukin;
 KW HLA class I gene; HLA-B*27; HLA-J; HLA-A2; acute phase protein response;
 KW syngene; immune response; IL-8; IL-16; diagnosis; gene therapy;
 KW transcription.
 OS Synthetic.
 PN W09606188-A1.
 PD 29-FEB-1996.
 PF 17-AUG-1995; U10523.
 PR 18-AUG-1994; US-292902.
 PR 15-AUG-1995; US-515190.
 PA (CYTO-) CYTOGEN CORP.
 PI Fowlkes DM, Rodwell JD;
 DR WPI: 96-151391/15.
 PT Synthetic random nucleotide sequences encoding ligand binding
 PT domains - identified by screening library of vectors or peptide(s)
 PT and useful for gene therapy and diagnosis
 PS Claim 70; Page 196; 224pp; English.
 CC The sequences given in R91378-414 represent peptides that bind to
 CC highly specific DNA binding domains (HSDB)'s. These sequences were
 CC tested for binding to the H2-kappa-B oligonucleotide which contains
 CC the NF-kappa-B binding site, and comprises the sequence given in T13579.
 CC These peptides may be used in a composition for diagnosis and gene
 CC therapy and for modifying the transcription or activity of a gene.
 SQ Sequence 28 AA;

Query Match 72.7%; Score 48; DB 18; Length 28;
 Best Local Similarity 57.1%; Pred. No. 1.59e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 13 rapgcwh 19
 QY 116 RHPSCWN 122

RESULT 5
 ID R42832 standard; Protein; 129 AA.
 AC R42832;
 DT 27-APR-1994 (first entry)
 DE Cytochrome P450 homologue cDNA insert pCGP773 product.
 KW Transgenic plants; altered petal colour; carnation.
 KW Dianthus caryophyllus cv. Laguna.
 PN W09320206-A.
 PD 14-OCT-1993.
 PF 25-MAR-1993; AU0127.
 PR 27-MAR-1992; AU-001538.

PR 07-JAN-1993; AU-006598.
 PA (ITFL-) INT FLOWER DEV PTY LTD.
 PI Cornish EC, Holton TA, Tanaka Y;
 DR WPI: 93-336914/42.
 DR N-PSDB: Q49431.
 PT Nucleic acid isolate encoding flavonoid-3'-hydroxylase - is used to
 PT create transgenic plants with altered petal colour
 PS Disclosure; Fig 10A-B; 86pp; English.
 CC The sequence is that encoded by the nucleotide sequence
 CC of a cytochrome P450 homologue cDNA insert isolated from pCGP773.
 CC It may be used as a probe for the isolation of putative carnation
 CC 3'-hydroxylase cDNA clones.
 SQ Sequence 129 AA;

Query Match 72.7%; Score 48; DB 8; Length 129;
 Best Local Similarity 66.7%; Pred. No. 1.59e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 87 rdpdcw 92
 | | | |
 QY 116 RHPSCW 121

RESULT 6
 ID R27204 standard; Protein; 1805 AA.

AC R27204;
 DT 15-FEB-1993 (first entry)
 DE Rat nestin.
 KW Intermediate filament; central nervous system; brain tumour;
 KW neurofilament.
 OS Rattus norvegicus.
 PN WO9214821-A.
 PD 03-SEP-1992.
 PF 21-FEB-1992; U01375.
 PR 22-FEB-1991; US-660412.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Lendahl U, McKay RDG;
 DR WPI: 92-316175/38.
 DR N-PSDB: Q28398.
 PT Diagnosis of pre-disposition to brain tumours - using DNA
 PT encoding nestin which distinguishes neural multi-potential stem
 PT cells from neuronal, glial and muscle cells
 PS Claim 3; Fig 1; 63pp; English.
 CC The amino acid sequence of the nestin gene which was deduced from
 CC the nucleotide sequence, suggests that nestin is a member of the
 CC intermediate filament protein family. The rat nestin amino acid
 CC sequence shows 75% similarity with the human nestin sequence. There
 CC is more than 60% identity between the two sequences. Antibodies to
 CC nestin protein can be used in vivo diagnosis of brain tumours.
 CC See also Q28399 for the human nestin gene.
 SQ Sequence 1805 AA;

Query Match 72.7%; Score 48; DB 5; Length 1805;
 Best Local Similarity 50.0%; Pred. No. 1.59e+02;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1588 qpacwd 1593
 | | | |
 QY 117 HPSCWN 122

RESULT 7

ID R60126 standard; Protein; 1805 AA.
 AC R60126;
 DT 21-MAR-1995 (first entry)
 DE Rat nestin protein is useful to identify brain tumours.
 KW nestin gene; brain tumour; neoplastic cells; glial; neuronal;
 KW muscle; neural multipotential stem cell; mammalian brain; detection;
 KW diagnosis; medulloblastoma; glioblastoma; oligodendroglioma.
 OS Rattus rattus.
 FH Key Location/Qualifiers
 FT misc_difference 167..168
 /note= "encoded by CACGG"

FT misc_difference 172..173
 /note= "encoded by CCG G GCC"
 PN US5338839-A.
 PD 16-AUG-1994.
 PF 12-APR-1988; 180548.
 PR 12-APR-1988; US-180548.
 PR 02-JUN-1988; US-201762.
 PR 25-OCT-1990; US-603803.
 PR 22-FEB-1991; US-660412.
 PR 19-MAR-1992; US-853913.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Lendahl U, McKay RDG;
 DR WPI: 94-263332/32.
 DR N-PSDB: Q70447.
 PT Nucleotide and protein sequences for human and rat nestin -
 PT distinguishes neural multipotential stem cells and brain tumour
 PT cells from more differentiated cell types; for use in the
 PT diagnosis of brain tumours
 PS Claim 4; Column 35-44; 45pp; English.
 CC This sequence is the rat nestin protein encoded by Q70447. Its
 CC degree of sequence homology to intermediate filaments in the core
 CC domain, its dissimilarity in the head and tail domains, and its
 CC different splicing pattern suggest that it defines a new class of
 CC intermediate filament protein. Nestin protein expression distinguishes
 CC neural multipotential stem cells and brain tumour cells from the more
 CC differentiated neural cell types (eg., neuronal, glial and muscle
 CC cells of the adult brain). The nestin protein can be used in diagnosing
 CC tumours of the brain, such as medulloblastomas, glioblastomas and
 CC oligodendroglioma. (See also R60127).
 SQ Sequence 1805 AA;

Query Match 72.7%; Score 48; DB 11; Length 1805;
 Best Local Similarity 50.0%; Pred. No. 1.59e+02;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1588 qpacwd 1593
 | | | |
 QY 117 HPSCWN 122

RESULT 8

ID W10423 standard; Protein; 586 AA.
 AC W10423;
 DT 17-OCT-1997 (first entry)
 DE Cytomegalovirus UL84 protein used for the treatment of infection.
 KW Cytomegalovirus; CMV; infection; prophylaxis; prevention; vaccine;
 KW immunisation; UL84.
 OS Human cytomegalovirus strain AD169.
 PN DL9527129-Al.
 PD 30-JAN-1997.
 PF 25-JUL-1995; 027129.
 PR 25-JUL-1995; DE-027129.
 PA (PLAC/) PLACHTER B.
 PA (STAM/) STAMMING T.
 PA (BEHW) BEHRINGERWERKE AG.
 PI Plachter B, Stammeringer T;
 DR WPI: 97-100836/10.
 DR N-PSDB: T47439.
 PT Medicament contg. cytomegalovirus UL84 protein - useful for
 PT treatment of human CMV infection
 PS Claim 2; Page 16-19; 30pp; English.
 CC T47439 is a genomic DNA sequence encoding the UL84 protein of human
 CC cytomegalovirus (CMV). The protein can be produced recombinantly and
 CC used to generate UL84 antibodies which are useful in the treatment
 CC and diagnosis of CMV infection. The protein and fragments and DNA
 CC encoding them are also useful for vaccination to prevent and/or treat
 CC CMV infection.
 SQ Sequence 586 AA;

Query Match 71.2%; Score 47; DB 23; Length 586;
 Best Local Similarity 57.1%; Pred. No. 2.00e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 306 rhalcwh 312
 QY 116 RHPSCWN 122

RESULT 9

ID R12603 standard; Protein: 95 AA.
 AC R12603;
 DT 02-SEP-1991 (first entry)
 DE SIB 121 intestinal mucin.
 KW SMUC; intestinal mucin; diagnosis; cancer.
 OS Homo sapiens.
 PN WO9108217-A.
 PD 13-JUN-1991.
 PF 04-DEC-1990; U07087.
 PR 05-DEC-1989; US-447140.
 PA (REGC) UNIV OF CALIFORNIA.
 PI Kim YS, Gum JR;
 DR WPI: 91-193147/26.
 DR N-PSDB; Q12249.
 PT Nucleic acid encoding human intestinal mucin peptide(s) - used
 for producing prods. for diagnosis and treatment of disorders
 PT such as cancer
 PS Disclosure; Fig 7(C): 57pp; English.
 CC The consensus SIB-sequence has 5 serine residues and 7 threonine
 residues, in contrast to the SMUC-type consensus sequence, contg.
 CC 0 and 14 residues. The SIB-type consensus sequence contains
 CC only 1 proline residue as compared with 5 for the SMUC-type consensus
 CC sequence. The SIB-type consensus also has its serine and threonine
 CC residues clustered in groups of two or three.
 CC See also Q12217-21 and Q12247-51.
 SQ Sequence 95 AA;

Query Match 69.7%; Score 46; DB 3; Length 95;
 Best Local Similarity 66.7%; Pred. No. 2.52e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 86 rdphcw 91
 QY 116 RHPSCW 121

RESULT 10

ID W11283 standard; Protein: 163 AA.
 AC W11283;
 DT 27-APR-1997 (first entry)
 DE Human breast specific gene BSG6 partial polypeptide.
 KW Breast specific gene 6; BSG6; breast cancer; metastasis;
 KW molecular marker; diagnosis; therapy; vaccine; antibody;
 KW inhibitor.
 OS Homo sapiens.
 FH Kcy Location/Qualifiers
 FT misc_difference 14
 FT /note= "residue 14 corresponds to an in-frame
 stop codon in the nucleotide sequence"
 FT misc_difference 36
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 40
 FT /note= "residue 40 corresponds to an in-frame
 stop codon in the nucleotide sequence"
 FT misc_difference 70
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 71
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 72
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 103
 FT /note= "residue 103 corresponds to an in-frame
 stop codon in the nucleotide sequence"

FT misc_difference 105
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 108
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 116
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 118
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 124
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 125
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 130
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 135
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 137
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 142
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 147
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 155
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 159
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT misc_difference 160
 FT /note= "variable site, owing to degeneracy of
 the nucleotide sequence"
 FT WO9702280-A1.
 PN 23-JAN-1997.
 PD 30-JUN-1995; U08295.
 PR 30-JUN-1995; WO-U08295.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Ji H, Rosen CA;
 DR WPI: 97-108908/10.
 DR N-PSDB; T51188.
 PT New breast specific genes and proteins - used for detecting and
 treating breast cancer and developing a breast cancer vaccine
 PS Claim 8; Fig 6; 82pp; English.
 CC Breast specific polypeptides (W11280-83) are respectively encoded
 by breast specific genes BSG2, BSG3, BSG4 and BSG6. Recombinant
 breast specific polypeptides can be produced in transformed host
 CC cells. They can be used as diagnostic markers for breast cancers
 CC and in methods for determining if breast cancer has metastasized.
 CC Antibodies raised against the polypeptides can be used to target
 CC cancer cells and as part of a breast cancer vaccine. Antagonists
 CC can be employed in the treatment of breast cancer.
 SQ Sequence 163 AA;

Query Match 69.7%; Score 46; DB 21; Length 163;
 Best Local Similarity 66.7%; Pred. No. 2.52e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 78 qnpscw 83
 QY 116 RHPSCW 121

RESULT 11

ID P91395 standard; peptide; 383 AA.
 AC P91395;
 DT 15-FEB-1990 (first entry)
 DE Peptide expressed by E.coli Bio F gene.
 KW E.coli; Bio F; biotin.
 OS Escherichia coli.
 PN GB221630-A.
 PD 11-OCT-1989.
 PF 17-MAR-1989; 006210.
 PR 22-MAR-1988; GB-006804.
 ER 17-MAR-1989; GB-006210.
 FA (URAG-) UK Min. Agric. Fish.
 PI Pearson BM, McKee RA;
 DR WPI: 89-295085/41.
 DR N-PSDB: N91333.
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 PS Table 6; page 39-40; 52pp; English.
 CC The enzyme is expressed by the E.coli bio F gene. It is a biotin
 CC synthetase enzyme, and permits the transformant to grow in the absence of
 CC exogenous biotin, giving a useful marker for yeast in fermentation.
 CC Sequence 383 AA;
 SQ

Query Match 68.2%; Score 45; DB 1; Length 383;
 Best Local Similarity 50.0%; Pred. No. 3.17e+02;
 Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 338 rqqgcw 343
 I: :|||
 QY 116 RHPSCW 121

RESULT 12
 ID R51884 standard; Protein; 384 AA.
 AC R51884;
 DT 17-NOV-1994 (first entry)
 DE KAPA synthase encoded by bioF gene in plasmid pB030A-15/9.
 KW Biotin; expression; enterobacteria; vitamin H; synthesis;
 KW plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioH;
 KW promoter; ptac; biotin synthase; KAPA synthase;
 KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
 KW dethiobiotin synthase; DAPA synthase;
 KW S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;
 KW seborrhea; dermatitis.
 OS Escherichia coli DSM498.
 PN W09408023-A.
 PD 14-APR-1994.
 PF 01-OCT-1993; E02688.
 PR 02-OCT-1992; CH-003124.
 PR 15-JUL-1993; CH-002134.
 FA (LONZ) LONZA AG.
 PI Birch O, Brass J, Fuhrmann M, Shaw N;
 DR WPI: 94-135587/16.
 DR N-PSDB: 062386.
 PT Biotechnological biotin prodn. using enterobacterial biotin-gene
 PT - providing vitamin H in high yield
 PS Disclosure; Fig 6, page 47-55 and 60-65; 92pp; German.
 CC Plasmid pB030A-15/9 contains the bioB, bioF, bioC, bioD and bioA
 CC genes responsible for biosynthesis of biotin, arranged in a
 CC transcription unit. Microorganisms contg. these DNA fragments or
 CC plasmids may be used in the prodn. of biotin. Biotin (Vitamin H)
 CC may prevent seborrhea, dermatitis, loss of appetite and tiredness.
 SQ Sequence 384 AA;

Query Match 68.2%; Score 45; DB 10; Length 384;
 Best Local Similarity 50.0%; Pred. No. 3.17e+02;
 Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 339 rqqgcw 344
 I: :|||
 QY 116 RHPSCW 121

RESULT 13
 ID R10274 standard; Protein; 502 AA.
 AC R10274;
 DT 04-APR-1991 (first entry)
 DE Simian immunodeficiency virus gag protein.
 KW siVmd; HIV; AIDS; vaccine; pSMH103.
 OS Simian immunodeficiency virus.
 PN J02299587-A.
 PD 11-DEC-1990.
 PF 11-MAY-1989; 116129.
 PR 11-MAY-1989; JP-116129.
 PA (TOFU) TONEN CORP.
 DR WPI: 91-027566/04.
 DR N-PSDB: Q10203.
 PT New complementary DNA to RNA gene of simian immuno-deficiency
 PT virus - used for preparing vaccine and diagnosis agent for AIDS
 PS Claim 2; Fig 4; 14pp; Japanese.
 CC SIV proteins may be produced from an E.coli expression system
 CC transformed with plasmid pSMH103.
 CC These may be used in diagnosis, treatment, and development of a vaccine
 CC against AIDS, as the sequence has the same structure as HIV-1 (but
 CC lacking the vpx gene).
 SQ Sequence 502 AA;

Query Match 68.2%; Score 45; DB 2; Length 502;
 Best Local Similarity 57.1%; Pred. No. 3.17e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 405 rrgcwn 411
 I: :|||
 QY 116 RHPSCW 122

RESULT 14
 ID W18305 standard; Protein; 1565 AA.
 AC W18305;
 DT 30-JAN-1998 (first entry)
 DE Photorhabdus luminescens insect toxin TccB.
 KW Insecticide; insect; toxin; pest control; biological control;
 KW Photorhabdus luminescens; TccB; Southern corn rootworm;
 KW Colorado potato beetle; Western corn rootworm; meal worm;
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
 KW Diptera, Dictyoptera; Acarina; Homoptera.
 OS Photorhabdus luminescens strain W-14 (ATCC 55397).
 FH Key Location/Qualifiers
 FT Peptide 1..11
 FT /note= "N-terminal peptide (Claim 30)"

PN W09717432-A1.
 PD 15-MAY-1997.
 PF 06-NOV-1996; U18003.
 PR 28-AUG-1996; US-705484.
 PR 06-NOV-1995; US-007255.
 PR 28-FEB-1996; US-608423.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;
 PI Firench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 PI Petelli J, Roberts JL, Rocheleau TA, Schoonover S;
 PI Strickland JA;
 DR WPI: 97-281022/25.
 DR N-PSDB: T68849.
 PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
 PT can be genetically engineered into insect larvae food and plants for
 PT insect control
 PS Claim 34; Page 239-245; 276pp; English.
 CC This polypeptide comprises the 185 kDa TccB insecticidal toxin
 CC protein of Photorhabdus luminescens W-14. Its sequence was deduced
 CC from a genomic DNA clone (T68849) and includes an isolated
 CC N-terminal peptide (see W17867). Claimed toxins of P. luminescens
 CC (see W17871, W17884-89, W17899-900, W18301-06) can be produced by
 CC recombinant DNA methods and applied to, or genetically engineered

CC Into, insect larvae food and plants for insect control. The toxins
CC are particularly effective against Southern corn rootworm, Colorado
CC potato beetle, Western corn rootworm, meal worm, boll weevil and
CC turf grub (Coleoptera), beet armyworm, black cutworm, cabbage looper,
CC codling moth, corn earworm, European corn borer, tobacco hornworm
CC and tobacco budworm (Lepidoptera), and are also active against
CC insects of the orders Hymenoptera, Diptera, Dictyoptera, Acarina
CC and Homoptera. (All claimed).

SQ Sequence 1565 AA;

Query Match 68.2%; Score 45; DB 25; Length 1565;

Best Local Similarity 80.0%; Pred. No. 3.17e+02;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 215 pncwn 219

|:|:|

Qy 118 PSCWN 122

RESULT 15

ID W56568 standard; Protein; 1565 AA.

AC W56568;

DT 07-AUG-1998 (first entry)

DE toxin TccB, encoded by the tccB gene from genomic region tcc.

KW Photorhabdus luminescens W-14; nematode; symbiotic;

KW Heterorhabditis; tca; tcb; tcc; tcd; insecticidal activity; toxin;

KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;

KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;

KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;

KW cabbage looper; codling moth; corn earworm; European corn borer;

KW Tobacco hornworm; budworm.

OS Photorhabdus luminescens.

PN WO9808932-A1.

PD 05-MAR-1998.

PF 05-MAY-1997; U07657.

PR 08-NOV-1996; WO-U18003.

PR 28-AUG-1996; US-705484.

PR 06-NOV-1996; US-743699.

PA (DWC) DOWELANCO.

PA (WISC) WISCONSIN ALUMNI RES FOUND.

PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R,

PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL,

PI Petell J, Roberts JL, Rocheleau TA, Schoonover S,

PI Strickland JA, Sukhapinda K;

DR WPI: 98-179427/16.

DR N-PSDB: V29926.

PT Isolated toxins from Photorhabdus luminescens strains - useful for

PT control of insect pests

PS Claim 34; Pages 276-281; 321pp; English.

CC The present sequence represents a protein named TccB of the bacterium

CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the

CC nematodes of the Heterorhabditis genus. The bacterium has at least 4

CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are

CC produced from these regions that are associated with insecticidal

CC activity. The native toxins are secreted proteins. The proteins are

CC toxic to insects upon exposure and especially when ingested. The

CC nucleic acid sequence can be used to produce transgenic plants,

CC baculoviruses or microbial hosts for toxin production. They can be used

CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,

CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the

CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,

CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,

CC codling moth, corn earworm, European corn borer or tobacco hornworm

CC or budworm.

SQ Sequence 1565 AA;

Query Match 68.2%; Score 45; DB 31; Length 1565;

Best Local Similarity 80.0%; Pred. No. 3.17e+02;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 215 pncwn 219

|:|:|

Qy 118 PSCWN 122

Search completed: Thu Jul 8 18:22:10 1999
Job time : 18 secs.

W P R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:20:36 1999; MasPar time 5.83 Seconds
48.098 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (116-122) from US09041236.pep (8 of 45)
Perfect Score: 66
Sequence: 1 RHPSCWN 7

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.145; Variance 30.037; scale 0.704

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description Pred. No.
1	53	80.3	327 1 RDVZAS ribonucleoside-diphosph 1.25e+00
2	53	80.3	343 2 T02932 cytochrome P-450 - ma 1.25e+00
3	53	80.3	530 2 T03259 cytochrome P450 - mai 1.25e+00
4	53	80.3	535 2 T03258 cytochrome P450 - mai 1.25e+00
5	51	77.3	402 2 C71440 hypothetical protein 3.05e+00
6	51	77.3	949 2 S34020 probable membrane pro 3.05e+00
7	51	77.3	1058 2 A48195 ubiquitin--protein li 3.05e+00
8	50	75.8	145 2 G59312 hypothetical protein 4.74e+00
9	50	75.8	312 1 J01559 polyhedron envelope p 4.74e+00
10	50	75.8	505 2 S38534 cytochrome P450 76A2 4.74e+00
11	50	75.8	587 1 WNBETE 65k early nonstructur 4.74e+00
12	48	72.7	1805 2 A34736 nestin - rat 1.13e+01
13	48	72.7	2175 1 GNNYBE genome polyprotein - 1.13e+01
14	47	71.2	219 1 R3KT3 ribosomal protein S3, 1.72e+01
15	47	71.2	434 2 S7331 hypothetical protein 1.72e+01
16	47	71.2	441 2 G64516 hypothetical protein 1.72e+01
17	47	71.2	534 2 S57974 hypothetical protein 1.72e+01
18	47	71.2	560 2 S50439 hypothetical protein 1.72e+01
19	47	71.2	586 1 WNBETE 65k early nonstructur 1.72e+01
20	47	71.2	609 2 JC61134 DNA-directed RNA poly 1.72e+01
21	46	69.7	36 2 E44336 neurotoxin Tx3-5 - sp 2.62e+01
22	46	69.7	71 2 S78193 ribosomal protein L31 2.62e+01
23	46	69.7	281 2 S56338 hypothetical 31.8k pr 2.62e+01

```
ORGANISM      #formal_name Zea mays #common_name maize
DATE          24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
ACCESSIONS    T02932
REFERENCE      Z14778
#authors      Sisco, P.H.
#submission   submitted to the EMBL Data Library, May 1997
#description   Agr(c94) random cDNA clone from Zea mays etiolated
               coleoptiles.
#accession    T02932
#status       preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-343 #label SIS
#cross-references EMBL:AF004210; NID:g2199565; PID:g2199566
GENETICS      CYP71C1
#gene         CYP71C1
#summary      #length 343 #checksum 9900

Query Match   80.3%; Score 53; DB 2; Length 343;
Best Local Similarity 57.1%; Pred. No. 1.25e+00;
Matches       4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 239 RDPTCWD 245
QY 116 RHPSCWN 122
|:|:|:

RESULT 3
ENTRY   #type complete
TITLE   cytochrome P450 - maize
ORGANISM #formal_name Zea mays #common_name maize
DATE    24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
ACCESSIONS T03259
REFERENCE   T03259
#authors    Frey, M.; Klem, R.; Saedler, H.; Gierl, A.
#journal    Mol. Gen. Genet. (1995) 246:100-109
#title      Expression of a cytochrome P450 gene family in maize.
#accession  T03259
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues   1-530 #label FRE
#cross-references EMBL:X81827; NID:g550435; PID:g550436
GENETICS    CYP71C1
#note      CYP71C1
#summary    #length 530 #molecular-weight 59282 #checksum 882

Query Match   80.3%; Score 53; DB 2; Length 530;
Best Local Similarity 57.1%; Pred. No. 1.25e+00;
Matches       4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 426 RDPTCWD 432
QY 116 RHPSCWN 122
|:|:|:

RESULT 4
ENTRY   #type complete
TITLE   cytochrome P450 - maize
ORGANISM #formal_name Zea mays #common_name maize
DATE    24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
ACCESSIONS T03258
REFERENCE   T03258
#authors    Frey, M.; Klem, R.; Saedler, H.; Gierl, A.
#journal    Mol. Gen. Genet. (1995) 246:100-109
#title      Expression of a cytochrome P450 gene family in maize.
#accession  T03258
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues   1-535 #label FRE
#cross-references EMBL:X81828; NID:g550433; PID:g550434
GENETICS
```

```
321/3
CYP71C1
#length 535 #molecular-weight 59715 #checksum 429

Query Match   80.3%; Score 53; DB 2; Length 535;
Best Local Similarity 57.1%; Pred. No. 1.25e+00;
Matches       4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 426 RDPTCWD 432
QY 116 RHPSCWN 122
|:|:|:

RESULT 5
ENTRY   #type complete
TITLE   hypothetical protein - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-eur
#summary    Cress
               Columbia
               03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change
               05-Dec-1998
               C71440
               A71400
               Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.;
               Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.;
               Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel,
               K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.;
               Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.;
               Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.;
               Lecharny, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.;
               Kavanagh, T.; Hempel, S.; Kottler, P.; Entian, K.D.; Rieger,
               M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.;
               James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka,
               A.; Voukelatou, E.; Milioni, D.; Hatzopoulos, P.;
               Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.;
               Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.;
               Rechman, S.; Ansoorge, W.; Cooke, R.; Berger, C.; Dolsen,
               M.; Voet, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.;
               Schueller, C.; Chalwatzis, N.
               #journal    Nature (1998) 391:485-488
               #title      Analysis of 1.9 Mb of contiguous sequence from chromosome 4
               of Arabidopsis thaliana.
               #cross-references MUID:98121113
               #accession  C71440
               ##status    preliminary; nucleic acid sequence not shown;
               ##molecule_type DNA
               ##residues  1-402 ##label BEV
               ##cross-references GB:297343; NID:g2245073; PID:g2245074
               GENETICS
               #map_position 4COP9-4G3845
               #length 402 #molecular-weight 45520 #checksum 732

Query Match   77.3%; Score 51; DB 2; Length 402;
Best Local Similarity 83.3%; Pred. No. 3.05e+00;
Matches       5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 11 RPPSCW 16
QY 116 RHPSCW 121
|:|:|:

RESULT 6
ENTRY   #type complete
TITLE   probable membrane protein YMR019w - yeast (Saccharomyces
               cerevisiae)
ALTERNATE_NAMES hypothetical protein YMR711.08
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
               06-Feb-1998
               #accession  S54020
               #cross-references S54014
               #summary    Lye, G.; Churcher, C.M.
```

```
#submission submitted to the EMBL Data Library, May 1995
#accession S54020
#molecule_type DNA
##residues 1-949 ##label LYX
##cross-references EMBL:249211; NID:G798922; PID:g798929; MIPS:YMR019w
##experimental_source strain AB972
GENETICS
#gene SCD:STB4
##cross-references SGD:S0004621; MIPS:YMR019w
#map_position 13R
CLASSIFICATION #superfamily unassigned GAL4-type zinc cluster proteins; GAL4
zinc binuclear cluster homology
transmembrane protein
KEYWORDS
18-34 #domain transmembrane #status predicted #label TML\
FEATURE #domain GAL4 zinc binuclear cluster homology #label
82-118 GAL4\
756-772 #domain transmembrane #status predicted #label TM2
SUMMARY #length 949 #molecular-weight 109824 #checksum 5627
Query Match 77.3%; Score 51; DB 2; Length 949;
Best Local Similarity 57.1%; Pred. No. 3.05e+00;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 531 ROTACWN 537
QY 116 RHPSCWN 122
RESULT 7
ENTRY A48195 #type complete
TITLE ubiquitin-protein ligase E1 homolog - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
21-Aug-1998
ACCESSIONS A48195
REFERENCE A48195
#authors Kok, K.; Hofstra, R.; Pilz, A.; van den Berg, A.; Terpstra,
P.; Buys, C.H.; Carritt, B.
#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:6071-6075
#title A gene in the chromosomal region 3p21 with greatly reduced
expression in lung cancer is similar to the gene for
ubiquitin-activating enzyme.
#cross-references MUID:193317626
#accession A48195
##status preliminary
#molecule_type nucleic acid
##residues 1-1058 ##label KOK
##experimental_source pre-B cells
#note sequence extracted from NCBI backbone (NCBIN:134901,
NCBIP:134902)
CLASSIFICATION #superfamily ubiquitin-activating enzyme E1
SUMMARY #length 1058 #molecular-weight 116484 #checksum 7203
Query Match 77.3%; Score 51; DB 2; Length 1058;
Best Local Similarity 66.7%; Pred. No. 3.05e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 112 HPTCWS 117
QY 117 HPSCWN 122
RESULT 8
ENTRY G69312 #type complete
TITLE hypothetical protein AF0503 - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Jun-1998
ACCESSIONS G69312
REFERENCE A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham,
D.E.; Kypides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
Glock, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession G69312
##status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
##residues 1-145 ##label KLE
##cross-references GB:AE001069; GB:AE000782; NID:g2689392; PID:g2650128;
TIGR:AF0503
SUMMARY #length 145 #molecular-weight 16721 #checksum 7087
Query Match 75.8%; Score 50; DB 2; Length 145;
Best Local Similarity 66.7%; Pred. No. 4.74e+00;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 117 HPECWD 122
QY 117 HPSCWN 122
RESULT 9
ENTRY JQ1559 #type complete
TITLE polyhedron envelope protein - Lymantria dispar nuclear
polyhedrosis virus
ORGANISM #formal_name Lymantria dispar nuclear polyhedrosis virus,
LMNPV
#note host Lymantria dispar (gypsy moth)
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
24-Oct-1997
ACCESSIONS JQ1559
REFERENCE P03339
#authors Bjornson, R.M.; Rohrmann, G.F.
#journal J. Gen. Virol. (1992) 73:1499-1504
#title Nucleotide sequence of the polyhedron envelope protein gene
region of the Lymantria dispar nuclear polyhedrosis virus.
#cross-references MUID:92300345
#accession JQ1559
##molecule_type DNA
##residues 1-312 ##label BJO
##cross-references GB:D37947; GB:D10836; NID:g532333; PID:d1007741;
PID:g221993
CLASSIFICATION #superfamily baculovirus polyhedral envelope protein
KEYWORDS envelope protein
SUMMARY #length 312 #molecular-weight 33827 #checksum 566
Query Match 75.8%; Score 50; DB 1; Length 312;
Best Local Similarity 71.4%; Pred. No. 4.74e+00;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 50 RHKQWN 56
QY 116 RHPSCWN 122
RESULT 10
ENTRY S38534 #type complete
TITLE cytochrome P450 76A2 - eggplant
ORGANISM #formal_name Solanum melongena #common_name eggplant,
aubergine
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
```

05-Mar-1999
S38534
#accessions
#reference
#authors
#journal
#title
#cross-references MUID:94032483
#accession
#status
#molecule_type mRNA
#residues 1-505 ##label TOG
#cross-references EMBL:X71657; NID:g415910; PID:g415911

GENETICS
#gene
#classification
#keywords
#feature
#summary

CYP76A2
#superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
chromoprotein; heme

#domain cytochrome P450 homology #label CYP
#binding_site heme iron (Cys) (axial ligand) #status
predicted
#length 505 #molecular-weight 57807 #checksum 6089

Query Match 75.8%; Score 50; DB 2; Length 505;
Best Local Similarity 57.1%; Pred. No. 4.74e+00;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 408 RPECWD 414
Qy 116 RHPSCWN 122

RESULT 11
ENTRY
TITLE
ORGANISM
#note
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:92114132
#accession
#molecule_type mRNA
#residues 1-587 ##label HEY
#cross-references GB:M81432; NID:g330541; PID:g330542
#classification #superfamily human cytomagalovirus 65K early nonstructural protein

leucine zipper; nonstructural protein

114-135
325-373
#region leucine zipper motif
#region leucine zipper motif
#length 587 #molecular-weight 65387 #checksum 6601

Query Match 75.8%; Score 50; DB 1; Length 587;
Best Local Similarity 57.1%; Pred. No. 4.74e+00;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 307 RHAVCWH 313
Qy 116 RHPSCWN 122

RESULT 12
ENTRY
TITLE
ORGANISM

A34736 #type complete
nestin - rat
#formal_name Rattus norvegicus #common_name Norway rat

13-Jul-1990 #sequence_revision 09-Oct-1992 #text_change
10-Sep-1997
A34736
#accessions
#reference
#authors
#journal
#title
#cross-references MUID:90150286
#accession
#status
#molecule_type DNA; mRNA
#residues 1-1805 ##label LEN
#cross-references GB:M34384; NID:g205663; PID:g205664
#note
the authors translated the codon CCG for residue 168 as
Gly, ATC for residue 169 as Ser, GGT for residue 171
as Val, and CCG for residue 172 as Arg

colled coil
#length 1805 #molecular-weight 198744 #checksum 3947

Query Match 72.7%; Score 48; DB 2; Length 1805;
Best Local Similarity 50.0%; Pred. No. 1.13e+01;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1588 QPACWD 1593
Qy 117 HPSCWN 122

RESULT 13
ENTRY
TITLE
CONTAINS
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:88117392
#accession
#molecule_type genomic RNA
#residues 1-2175 ##label EAR
#cross-references GB:D00214; NID:g2696866; PID:d1024909
#classification #superfamily poliovirus genome polyprotein
coat protein; core protein; genome-linked protein;
nucleotidyltransferase; phosphoprotein; proteinase

#product coat protein VP4 #status predicted #label VP4
#product coat protein VP2 #status predicted #label VP2
#product coat protein VP3 #status predicted #label VP3
#product coat protein VP1 #status predicted #label VP1
#product core protein P2-A #status predicted #label P2A
#product core protein P2-B #status predicted #label P2B
#product core protein P2-C #status predicted #label P2C
#product core protein P3-A #status predicted #label P3A
#product genome-linked protein VPg #status predicted
#label VPg
#product proteinase #status predicted #label PTS
#product RNA-directed RNA polymerase #status predicted
#label RPS
#binding_site phosphoryl-RNA (Tyr) (covalent) #status
predicted
#length 2175 #molecular-weight 242502 #checksum 7260

Query Match 72.7%; Score 48; DB 1; Length 2175;
Best Local Similarity 66.7%; Pred. No. 1.13e+01;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 800 KHTSCW 805
:|||||

Qy 116 RHPSCW 121

RESULT 14

ENTRY R3KT3 #type complete
TITLE ribosomal protein S3, cyanelle - Cyanophora paradoxa cyanelle
ORGANISM #formal_name cyanelle Cyanophora paradoxa
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
30-Jan-1998

ACCESSIONS

REFERENCE S12212

#authors

Michalowski, C.B.; Pfanzagl, B.; Loeffelhardt, W.; Bohnert,

H.J.

#journal

Mol. Gen. Genet. (1990) 224:222-231

#title

The cyanelle S10 spc ribosomal protein gene operon from

Cyanophora paradoxa.

#cross-references

MUID:91117189

#accession

S12212

#molecule_type

DNA

#residues

1-219 #label MIC

##cross-references GB:M30487; NID:g336645; PID:g336647

##experimental_source strain LB555 UTEX

##note the authors translated the initiation codon GUG for

residue 1 as Val

GENETICS

#gene

rpS3

#genome

cyanelle

#start_codon

GUG

CLASSIFICATION #superfamily Escherichia coli ribosomal protein S3

KEYWORDS cyanelle; protein biosynthesis; ribosome

SUMMARY #length 219 #molecular-weight 25014 #checksum 7062

Query Match

Best Local Similarity 71.2%; Score 47; DB 1; Length 219;

Pred. No. 1.72e+01;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 17 KHRSCW 22

:|||||

Qy 116 RHPSCW 121

RESULT 15

ENTRY S73331 #type complete
TITLE hypothetical protein E07_orf434 - Mycoplasma pneumoniae (ATCC
29342) (SGC3)

ORGANISM #formal_name Mycoplasma pneumoniae

ATCC 29342

DATE 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change

17-Jul-1998

ACCESSIONS

REFERENCE S73331

S73327

#authors

Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li,

B.C.; Herrmann, R.

Nucleic Acids Res. (1996) 24:4420-4449

#journal Complete sequence analysis of the genome of the bacterium

#title Mycoplasma pneumoniae.

#cross-references MUID:97105885

#accession

S73331

##status preliminary; nucleic acid sequence not shown;

translation not shown

#molecule_type

DNA

##residues 1-434 #label HIM

##cross-references EMBL:AE000001; GB:U00089; NID:g1673645; PID:g1673650

##note the nucleotide sequence was submitted to the EMBL Data

Library, November 1996

GENETICS

#genetic_code SGC3

SUMMARY #length 434 #molecular-weight 47970 #checksum 671

Query Match 71.2%; Score 47; DB 2; Length 434;
Best Local Similarity 66.7%; Pred. No. 1.72e+01;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 330 RYPCW 335

:|||||

Qy 116 RHPSCW 121

Search completed: Thu Jul 8 18:20:46 1999

Job time : 10 secs.

THIS PAGE BLANK (USPTO)

WQESRCH

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:19:35 1999; MasPar time 3.37 Seconds
Tabular output not generated. 58.789 Million cell updates/sec

Title: >US-09-041-236-2
Description: (116-122) from US09041236.pep (8 of 45)
Perfect Score: 66
Sequence: 1 RHPSCWN 7

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 21.794; Variance 27.867; scale 0.782

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	53	80.3	327	1	RIR2_ASFM2	4.69e-01
2	53	80.3	334	1	RIR2_ASFB7	4.69e-01
3	53	80.3	535	1	C7C1_MAIZE	4.69e-01
4	51	77.3	949	1	YMP9_YEAST	1.24e+00
5	51	77.3	1011	1	UBAL_HUMAN	1.24e+00
6	50	75.8	312	1	VPHE_NPVLD	1.99e+00
7	50	75.8	505	1	C762_SOLME	1.99e+00
8	50	75.8	587	1	UL84_HCMVT	1.99e+00
9	49	74.2	262	1	Y069_NPVAC	3.19e+00
10	48	72.7	901	1	CR14_MAIZE	5.07e+00
11	48	72.7	1805	1	NEST_RAT	5.07e+00
12	48	72.7	2175	1	POLG_BOVEV	5.07e+00
13	47	71.2	219	1	RR3_CYPAPA	8.02e+00
14	47	71.2	441	1	Y007_METJA	8.02e+00
15	47	71.2	560	1	YEC0_YEAST	8.02e+00
16	47	71.2	586	1	UL84_HCMVA	8.02e+00
17	46	69.7	281	1	IYCL_ECOLI	1.26e+01
18	46	69.7	531	1	KPKY_EIMTE	1.26e+01
19	46	69.7	552	1	SYL_SULAC	1.26e+01
20	46	69.7	581	1	AMV1_SCHPO	1.26e+01
21	46	69.7	2109	1	RRPL_VSVJH	1.26e+01
22	46	69.7	2109	1	RRPL_VSVJO	1.26e+01
23	46	69.7	2127	1	RRPL_RABVS	1.26e+01

24	46	69.7	2142	1	RRPL_RABVP	RNA POLYMERASE BETA SU	1.26e+01
25	45	68.2	224	1	EFA4_MOUSE	EPHRIN-A4 PRECURSOR (E	1.97e+01
26	45	68.2	311	1	YDAO_ECOLI	HYPOTHETICAL 35.6 KD P	1.97e+01
27	45	68.2	312	1	YG37_YEAST	HYPOTHETICAL 35.7 KD P	1.97e+01
28	45	68.2	316	1	HXSE_PSEAE	HOMOSERINE KINASE (EC	1.97e+01
29	45	68.2	326	1	HX44_MOUSE	HOMOSERINE KINASE HX-A	1.97e+01
30	45	68.2	331	1	KPRS_SYN7	RIBOSE-PHOSPHATE PYROP	1.97e+01
31	45	68.2	384	1	BIOF_ECOLI	8-AMINO-7-OXONONANOATE	1.97e+01
32	45	68.2	502	1	GAG_SIVGB	GAG POLYPROTEIN [CONTA	1.97e+01
33	45	68.2	546	1	PUR2_AGABI	DELTA-1-PYRROLINE-5-CA	1.97e+01
34	45	68.2	574	1	ENV_SRV2	ENV POLYPROTEIN (COAT	1.97e+01
35	45	68.2	630	1	GIDA_PSEPU	GLUCOSE INHIBITED DIVI	1.97e+01
36	45	68.2	1165	1	CYA6_MOUSE	ADENYLATE CYCLASE, TYP	1.97e+01
37	45	68.2	1166	1	CYA6_RAT	ADENYLATE CYCLASE, TYP	1.97e+01
38	45	68.2	1194	1	DPOL_VZVD	RNA POLYMERASE (EC 2.7	1.97e+01
39	45	68.2	1302	1	RRPL_BT10	DNA DIRECTED RNA POLYM	1.97e+01
40	45	68.2	1309	1	YAB2_SCHPO	HYPOTHETICAL 150.5 KD	1.97e+01
41	45	68.2	2016	1	CIN5_HUMAN	SODIUM CHANNEL PROTEIN	1.97e+01
42	45	68.2	2019	1	CIN5_RAT	SODIUM CHANNEL PROTEIN	1.97e+01
43	45	68.2	2205	1	POIN_RUBVT	NONSTRUCTURAL POLYPROT	1.97e+01
44	44	66.7	323	1	AMFR_HUMAN	AUTOCRINE MOTILITY FAC	3.05e+01
45	44	66.7	510	1	GAG_SIVMS	GAG POLYPROTEIN [CONTA	3.05e+01

ALIGNMENTS

RESULT 1	RIR2_ASFM2	STANDARD:	PRT:	327 AA.
AC	P26713:			
DT	01-AUG-1992 (REL. 23, CREATED)			
DT	01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)			
DT	01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)			
DE	RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (EC 1.17.4.1)			
DE	(RIBONUCLEOTIDE REDUCTASE)			
OS	AFRICAN SWINE FEVER VIRUS (ISOLATE MALAWI LIL 20/1) (ASFV).			
OC	VIROSES; DSDNA VIRUSES, NO RNA STAGE;			
OC	AFRICAN SWINE FEVER-LIKE VIRUSES.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 91335775.			
RA	BOURNSNELL M., SHAW K., YANEZ R.J., VINUELA E., DIXON L.;			
RT	"The sequences of the ribonucleotide reductase genes from African			
RT	swine fever virus show considerable homology with those of the			
RT	orthopoxvirus, vaccinia virus.;"			
RL	VIROLOGY 184:411-416(1991).			
CC	-!- FUNCTION: PROVIDES: 2'DEoxyRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED			
CC	-!- CATALYTIC ACTIVITY: H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED			
CC	THIOREDOXIN.			
CC	-!- COFACTOR: CONTAINS TWO IRON IONS.			
CC	-!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.			
CC	-!- SUBUNIT: HETEROIMER OF A LARGE AND A SMALL CHAIN.			
CC	-!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE			
CC	SMALL CHAIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; M64728; ; NOT_ANNOTATED_CDS.			
DR	PIR; B40568; RDVZAS.			
DR	PROSITE; PS00368; RIBORED_SMALL; 1.			
DR	PFAM; PF00268; ribonuc_red; 1.			
DR	HSSP; P11157; 1XSM.			
KW	OXIDOREDUCTASE; DNA REPLICATION; IRON.			
FT	METAL 70 70			
FT	IRON 1 (BY SIMILARITY).			
FT	METAL 101 101			
FT	IRON 1 AND 2 (BY SIMILARITY).			
FT	METAL 104 104			
FT	IRON 1 (BY SIMILARITY).			
FT	METAL 164 164			
FT	IRON 2 (BY SIMILARITY).			

```
FT METAL 198 198 IRON 2 (BY SIMILARITY).
FT METAL 201 201 IRON 2 (BY SIMILARITY).
FT ACT_SITE 108 108 BY SIMILARITY.
SQ SEQUENCE 327 AA; 38966 MW; 373250F8 CRC32;

Query Match 80.3%; Score 53; DB 1; Length 327;
Best Local Similarity 71.4%; Pred. No. 4.69e-01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 18 OHPECWN 24
:| | | |
Qy 116 RHPSCWN 122

RESULT 2
ID RIR2_ASFB7 STANDARD; PRT; 334 AA.
AC P42492;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOTIDE REDUCTASE).
GN F334L.
OS AFRICAN SWINE FEVER VIRUS (STRAIN BA71V) (ASFV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE;
OC AFRICAN SWINE FEVER-LIKE VIRUSES.
CC [1]
RN SEQUENCE FROM N.A.
RA YANEZ R.J., RODRIGUEZ J.M., NOGAL M.L., YUSTE L., ENRIQUEZ C.,
RA RODRIGUEZ J.F., VINUELA E.;
RT "Analysis of the complete nucleotide sequence of African swine fever
RT virus.";
RL VIROLOGY 208:249-278(1995).
CC -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: 2'DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC THIOREDOXIN + H(2)O -> RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDOXIN.
CC -!- COFACTOR: CONTAINS TWO IRON IONS (BY SIMILARITY).
CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC SMALL CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U18466; G780414; -.
CC DR PROSITE: PS00368; RIBORED_SMALL; 1.
CC DR PFAM: PF00268; ribonuc_red; 1.
CC DR HSP: P1157; 1XSM.
CC KW OXIDOREDUCTASE; DNA REPLICATION; IRON.
FT METAL 77 77 IRON 1 (BY SIMILARITY).
FT METAL 108 108 IRON 1 AND 2 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 171 171 IRON 2 (BY SIMILARITY).
FT METAL 205 205 IRON 2 (BY SIMILARITY).
FT METAL 208 208 IRON 2 (BY SIMILARITY).
FT ACT_SITE 115 115 BY SIMILARITY.
SQ SEQUENCE 334 AA; 39806 MW; 7167C714 CRC32;

Query Match 80.3%; Score 53; DB 1; Length 334;
Best Local Similarity 71.4%; Pred. No. 4.69e-01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 25 OHPECWN 31
:| | | |
Qy 116 RHPSCWN 122

FT METAL 198 198 IRON 2 (BY SIMILARITY).
FT METAL 201 201 IRON 2 (BY SIMILARITY).
FT ACT_SITE 108 108 BY SIMILARITY.
SQ SEQUENCE 327 AA; 38966 MW; 373250F8 CRC32;

Query Match 80.3%; Score 53; DB 1; Length 327;
Best Local Similarity 71.4%; Pred. No. 4.69e-01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 18 OHPECWN 24
:| | | |
Qy 116 RHPSCWN 122

RESULT 2
ID RIR2_ASFB7 STANDARD; PRT; 334 AA.
AC P42492;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOTIDE REDUCTASE).
GN F334L.
OS AFRICAN SWINE FEVER VIRUS (STRAIN BA71V) (ASFV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE;
OC AFRICAN SWINE FEVER-LIKE VIRUSES.
CC [1]
RN SEQUENCE FROM N.A.
RA YANEZ R.J., RODRIGUEZ J.M., NOGAL M.L., YUSTE L., ENRIQUEZ C.,
RA RODRIGUEZ J.F., VINUELA E.;
RT "Analysis of the complete nucleotide sequence of African swine fever
RT virus.";
RL VIROLOGY 208:249-278(1995).
CC -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: 2'DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC THIOREDOXIN + H(2)O -> RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDOXIN.
CC -!- COFACTOR: CONTAINS TWO IRON IONS (BY SIMILARITY).
CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC SMALL CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U18466; G780414; -.
CC DR PROSITE: PS00368; RIBORED_SMALL; 1.
CC DR PFAM: PF00268; ribonuc_red; 1.
CC DR HSP: P1157; 1XSM.
CC KW OXIDOREDUCTASE; DNA REPLICATION; IRON.
FT METAL 77 77 IRON 1 (BY SIMILARITY).
FT METAL 108 108 IRON 1 AND 2 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 171 171 IRON 2 (BY SIMILARITY).
FT METAL 205 205 IRON 2 (BY SIMILARITY).
FT METAL 208 208 IRON 2 (BY SIMILARITY).
FT ACT_SITE 115 115 BY SIMILARITY.
SQ SEQUENCE 334 AA; 39806 MW; 7167C714 CRC32;

Query Match 80.3%; Score 53; DB 1; Length 334;
Best Local Similarity 71.4%; Pred. No. 4.69e-01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 25 OHPECWN 31
:| | | |
Qy 116 RHPSCWN 122

RESULT 3
ID C7C1_MAIZE STANDARD; PRT; 535 AA.
AC Q43250; Q43254;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 71C1 (EC 1.14.-.-).
GN CYP71C1.
OS ZEA MAYS (MAIZE).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; ZEA.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. CI31A;
RX MEDLINE; 95124291.
RA FREY M., KLIEM R., SAEDLER H., GIERL A.;
RT "Expression of a cytochrome P450 gene family in maize.";
RL MOL. GEN. GENET. 246:100-109(1995).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X81828; G550434; -.
CC DR EMBL: X81827; G550436; -.
CC DR PROSITE: PS00086; CYTOCHROME_P450; 1.
CC DR PFAM: PF00067; P450; 1.
CC KW OXIDOREDUCTASE; MONOOXYGENASE; TRANSMEMBRANE; HEME.
FT TRANSMEM 14 34 POTENTIAL.
FT BINDING 470 470 HEME (BY SIMILARITY).
FT CONFLICT 6 6 G -> A (IN G550436).
FT CONFLICT 138 138 S -> C (IN G550436).
FT CONFLICT 148 148 V -> E (IN G550436).
FT CONFLICT 268 268 S -> G (IN G550436).
SQ SEQUENCE 535 AA; 59715 MW; 3236F5BB CRC32;

Query Match 80.3%; Score 53; DB 1; Length 535;
Best Local Similarity 57.1%; Pred. No. 4.69e-01;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 426 RDPICWD 432
| | | |
Qy 116 RHPSCWN 122

RESULT 4
ID YMP9_YEAST STANDARD; PRT; 949 AA.
AC P50104;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PUTATIVE 109.8 KD TRANSCRIPTIONAL REGULATORY PROTEIN IN SOK2-FMS1
DE INTERGENIC REGION.
GN YMR019W OR YMR711.08.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RA LYE G., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC -----
```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: 249211; G798929; -
DR PROSITE: PS00463; ZN2_Cy6_FUNGAL_1; 1.
DR PROSITE: PS00463; ZN2_Cy6_FUNGAL_2; 1.
DR PFAM: PF00172; ZN2_Clus; 1.
KW HYPOTHETICAL PROTEIN; TRANSCRIPTION REGULATION; DNA-BINDING;
KW NUCLEAR PROTEIN; ZINC; METAL-BINDING.
FT DNA_BIND 87 113 ZN(2)-CY5(6); FUNGAL-TYPE.
SQ SEQUENCE 949 AA; 109824 MW; 96BD0416 CRC32;

Query Match 77.3%; Score 51; DB 1; Length 949;
Best Local Similarity 57.1%; Pred. No. 1.24e+00;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 531 ROTACWN 537
|: |||
QY 116 RHPSCWN 122

RESULT 5
ID UBAL_HUMAN STANDARD; PRT; 1011 AA.
AC P41226;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE UBIQUITIN-ACTIVATING ENZYME E1 HOMOLOG (D8).
GN UBE1L OR UBE2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=B-CELL;
RX MEDLINE; 93317626.
RA KOK K., HOFSTRA R., PILZ A., VAN DEN BERG A., TERPSTRA P.,
RA BUYS C.H.C.M., CARRIET B.
RT "A gene in the chromosomal region 3p21 with greatly reduced
RT expression in lung cancer is similar to the gene for
RT ubiquitin-activating enzyme."
RL PROC. NATL. ACAD. SCI. U.S.A. 90:6071-6075(1993).
CC -!- FUNCTION: ACTIVATES UBIQUITIN BY FIRST ADENYLATING WITH ATP ITS
CC CARBOXY-TERMINAL GLYCINE RESIDUE AND THEREAFTER LINKING THIS
CC RESIDUE TO THE SIDE CHAIN OF A CYSTEINE RESIDUE IN E1, YIELDING
CC AN UBIQUITIN-E1 THIOLESTER AND FREE AMP.
CC -!- PATHWAY: FIRST STEP IN UBIQUITIN CONJUGATION.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF NORMAL AND TUMOR
CC CELL TYPES, BUT IS REDUCED IN LUNG CANCER CELL LINES.
CC -!- THERE ARE TWO ACTIVE SITES WITHIN THE E1 MOLECULE, ALLOWING IT TO
CC ACCOMMODATE TWO UBIQUITIN MOIETIES AT A TIME, WITH A NEW UBIQUITIN
CC FORMING AN ADENYLATE INTERMEDIATE AS THE PREVIOUS ONE IS
CC TRANSFERRED TO THE THIOLE SITE.
CC -!- SIMILARITY: SOME, E.COLI MOEB/CHLN, THIF AND ANABAENA SP. HESA.
CC -!- SIMILARITY: TO E1 IN OTHER SPECIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L13852; G986881; -
DR EMBL: L34170; -; NOT_ANNOTATED_CDS.
DR MIM, 191325; -;

DR PROSITE: PS00536; UBIQUITIN_ACTIVAT_1; FALSE_NEG.
DR PROSITE: PS00865; UBIQUITIN_ACTIVAT_2; 1.
DR PFAM: PF00899; Thif_family; 2.
KW UBIQUITIN CONJUGATION; LIGASE; MULTIGENE FAMILY; REPEAT.
FT ACT_SITE 23 598 BY SIMILARITY.
FT DOMAIN 53 574 2 APPROXIMATE REPEATS.
FT REPEAT 23 159 1-1.
FT REPEAT 422 574 1-2.
FT SIMILAR 461 565 TO E.COLI THIF (AA 53-153).
SQ SEQUENCE 1011 AA; 111719 MW; D7F82157 CRC32;

Query Match 77.3%; Score 51; DB 1; Length 1011;
Best Local Similarity 66.7%; Pred. No. 1.24e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 65 HPTCWS 70
|: |||
QY 117 HPCSWN 122

RESULT 6
ID VPHE_NPVLD STANDARD; PRT; 312 AA.
AC P36865;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE POLYHEDRAL ENVELOPE PROTEIN (PE) (POLYHEDRAL CALYX PROTEIN).
OS LYMANTRIA DISPAR MULTICAPSID NUCLEAR POLYHEDROSIS VIRUS (LDMNPV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
OC NUCLEOPOLYHEDROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92300345.
RA BJORNSEN R.M., ROHRMANN G.F.;
RT "Nucleotide sequence of the polyhedron envelope protein gene region
RT of the Lymantria dispar nuclear polyhedrosis virus."
RL J. GEN. VIROL. 73:1499-1504(1992).
CC -!- FUNCTION: MAJOR COMPONENT OF THE POLYHEDRA ENVELOPE.
CC -!- SIMILARITY: TO PE FROM OTHER BACULOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D37947; G221993; -
DR FIR: J01559; J01559.
KW ENVELOPE PROTEIN; VIRAL OCCLUSION BODY.
FT DOMAIN 60 68 POLY-GLY.
FT DOMAIN 301 309 POLY-ALA.
SQ SEQUENCE 312 AA; 33827 MW; 47FB0D1F CRC32;

Query Match 75.8%; Score 50; DB 1; Length 312;
Best Local Similarity 71.4%; Pred. No. 1.99e+00;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 50 RHKKCN 56
|: |||
QY 116 RHPSCWN 122

RESULT 7
ID C762_SOLME STANDARD; PRT; 505 AA.
AC P37122;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 76A2 (EC 1.14.-.-) (CYPLXXVIA2) (P-450EG7).
GN CYP76A2 OR CYP67.
OS SOLANUM MELONGENA (EGGPLANT) (AUBERGINE).

CC EURARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
CC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
CC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANACEAE; SOLANUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, SINSADHARANASU; TISSUE=HYPOCOTYL;
RX MEDLINE; 94032483.
RA TOGURI T., KOBAYASHI O., UMEMOTO N.;
RT "The cloning of eggplant seedling cDNAs encoding proteins from a
RT novel cytochrome P-450 family (Cyp76).";
RL BIOCHIM. BIOPHYS. ACTA 1216:165-169(1993).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X71657; G415911; -.
DR PIR; S38534; S38534.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PFAM; PF00067; P450; 1.
KW OXIDOREDUCTASE; MONOOXYGENASE; HEME; MULTIGENE FAMILY.
FT DOMAIN 36 43 PRO-RICH.
FT BINDING 448 448 HEME (BY SIMILARITY).
SQ SEQUENCE 505 AA; 57807 MW; 98EF7421 CRC32;

Query Match 75.8%; Score 50; DB 1; Length 505;
Best Local Similarity 57.1%; Pred. No. 1.99e+00;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 408 RDPCCWD 414
Qy 116 RHPSCWN 122
| | | |
| | | |

RESULT 8 STANDARD; PRT; 587 AA.
ID UL84_HCMVT
AC P29839;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 65 KD EARLY NONSTRUCTURAL PROTEIN (UL84 PROTEIN).
GN UL84.
OS HUMAN CYTOMEGALOVIRUS (STRAIN TOWNE).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92114132.
RA HE Y.S., XU L., HUANG E.S.;
RT "Characterization of human cytomegalovirus UL84 early gene and
RT identification of its putative protein product.";
RL J. VIROL. 66:1098-1108(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M81432; G330542; -.
DR PIR; A41808; WMBETE.
DR NONSTRUCTURAL PROTEIN.
KW NONSTRUCTURAL PROTEIN.
FT DOMAIN 9 19 ARG-RICH (BASIC).
FT DOMAIN 162 170 LYS-RICH (BASIC).
FT DOMAIN 171 183 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 587 AA; 65388 MW; ACE663AE CRC32;

Query Match 75.8%; Score 50; DB 1; Length 587;
Best Local Similarity 57.1%; Pred. No. 1.99e+00;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 307 RHAVCWH 313
Qy 116 RHPSCWN 122
| | | |
| | | |

RESULT 9 STANDARD; PRT; 262 AA.
ID Y069_NPVAC
AC P41469;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 30.4 KD PROTEIN IN LER3-IAP2 INTERGENIC REGION.
OS AUTOGRAHA CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS (ACMNPV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
OC NUCLEOPOLYHEDROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE; 94303173.
RA AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERBER M., POSSEE R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus.";
RL VIROLOGY 202:586-605(1994).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L22858; G559138; -.
DR HYPOTHETICAL PROTEIN.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 262 AA; 30355 MW; 30C0E870 CRC32;

Query Match 74.2%; Score 49; DB 1; Length 262;
Best Local Similarity 66.7%; Pred. No. 3.19e+00;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 35 RRPCCW 40
Qy 116 RHPSCW 121
| | | |
| | | |

RESULT 10 STANDARD; PRT; 901 AA.
ID CRI4_MAIZE
AC O24585;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PUTATIVE RECEPTOR PROTEIN KINASE CRINKLY4 PRECURSOR (EC 2.7.1.1-).
GN CRA.
OS ZEA MAYS (MAIZE).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; ZEA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. B73;
RX MEDLINE; 96355669.
RA BECRAFT P.W., STINARD P.S., MCCARTY D.R.;
RT "CRINKLY4: A TNFR-like receptor kinase involved in maize epidermal
RT differentiation.";
RL SCIENCE 273:1406-1409(1996).
CC -1- FUNCTION: PUTATIVE RECEPTOR PROTEIN KINASE. COULD PLAY A ROLE IN A
CC DIFFERENTIATION SIGNAL. THE CRINKLY4 (CRA) MUTATION AFFECTS LEAF
CC EPIDERMIS DIFFERENTIATION SUCH THAT CELL SIZE AND MORPHOLOGY ARE

```
CC ALTERED, AND SURFACE FUNCTIONS ARE COMPROMISED, ALLOWING GRAFT-
CC LIKE FUSIONS BETWEEN ORGANS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN
CC -!- SIMILARITY: BELONGS TO THE SER/THR-PROTEIN KINASE FAMILY.
CC -!- SIMILARITY: CONTAINS A PARTIAL LA-NGFR/TNFR-TYPE CYSTEINE-RICH
CC REGION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67422; G1597723; -
CC MAIZEDB; 128723; -
CC DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC DR PROSITE; PS00652; TNFR_NGFR_1; FALSE_NEG.
CC DR PROSITE; PS30050; TNFR_NGFR_2; 1.
CC DR PFAM; PF00020; TNFR_C6; 1.
CC DR PFAM; PF00069; pkinase; 1.
CC DR RECEPTOR; TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
CC REPEAT; TRANSMEMBRANE; SIGNAL.
CC FT SIGNAL 1 24
CC FT CHAIN 25 901
CC -----
CC DOMAIN 25 423
CC FT TRANSMEM 424 444
CC FT DOMAIN 445 901
CC FT DOMAIN 33 330
CC FT REPEAT 33 68
CC FT REPEAT 72 107
CC FT REPEAT 125 160
CC FT REPEAT 162 195
CC FT REPEAT 203 236
CC FT REPEAT 253 287
CC FT REPEAT 292 330
CC FT REPEAT 357 391
CC FT DOMAIN 505 712
CC FT NP_BIND 511 519
CC FT BINDING 533 533
CC FT ACT_SITE 634 634
CC FT ACT_SITE 901 AA; 97438 MW; 90597240 CRC32;
CC -----
CC Query Match 72.7%; Score 48; DB 1; Length 901;
CC Best Local Similarity 66.7%; Pred. No. 5.07e+00;
CC Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC -----
CC Db 90 ROPYCW 95
CC QY 116 RHPSCW 121
CC -----
CC RESULT 11
CC ID NEST-RAT STANDARD; PRT; 1805 AA.
CC AC P21263;
CC DT 01-MAY-1991 (REL. 18, CREATED)
CC DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
CC DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
CC DE NESTIN.
CC GN NES.
CC OS RATTUS NORVEGICUS (RAT).
CC OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
CC OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 90150286.
CC RA LENDAHU U., ZIMMERMAN L.B., MCKAY R.D.G.;
CC "CNS stem cells express a new class of intermediate filament
CC protein.";
```

```
RL CELL 60:585-595(1990).
CC -!- TISSUE SPECIFICITY: CNS STEM CELLS.
CC -!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
CC IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M34384; G205664; -
CC PIR; A34736; A34736.
CC DR PROSITE; PS00226; IF; 1.
CC DR PFAM; PF00038; filament; 2.
CC KW INTERMEDIATE FILAMENT; COILED COIL; HEPTAD REPEAT PATTERN;
CC NEURONE.
CC FT DOMAIN 1 7
CC FT DOMAIN 8 314
CC FT DOMAIN 315 1805
CC FT DOMAIN 8 43
CC FT DOMAIN 44 55
CC FT DOMAIN 56 151
CC FT DOMAIN 152 174
CC FT DOMAIN 175 193
CC FT DOMAIN 194 196
CC FT DOMAIN 197 314
CC FT SEQUENCE 1805 AA; 198744 MW; 789C2B48 CRC32;
CC -----
CC Query Match 72.7%; Score 48; DB 1; Length 1805;
CC Best Local Similarity 50.0%; Pred. No. 5.07e+00;
CC Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC Db 1588 QPACWD 1593
CC QY 117 HPSCWN 122
CC -----
CC RESULT 12
CC ID POLG-BOVEV STANDARD; PRT; 2175 AA.
CC AC P12915;
CC DT 01-OCT-1989 (REL. 12, CREATED)
CC DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
CC DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
CC DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
CC P2A TO P2C, P3A; GENOME-LINKED PROTEIN VP3; PICORNAIN 3C
CC DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
CC DE (EC 2.7.7.48)].
CC OS BOVINE ENTEROVIRUS (STRAIN VG-5-27) (BEV).
CC OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
CC OC ENTEROVIRUS.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 88117392.
CC RA EARLE J.A.P., SKUCE R.A., FLEMING C.S., HOEY E.M., MARTIN S.J.;
CC "The complete nucleotide sequence of a bovine enterovirus.";
CC J. GEN. VIROL. 69:253-263(1988).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS.
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```



```

DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DN HYPOTHETICAL 61.3 KD PROTEIN IN URA3-MMS21 INTERGENIC REGION.
GN YEL020C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGEN R., OEFNER P., OH C.,
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- COFACTOR: THIAMINE PYROPHOSPHATE (BY SIMILARITY).
CC -!- SIMILARITY: STRONG. TO BACTERIAL OXALYL-COA DECARBOXYLASE.
CC -!- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U18530; G602387; -
DR PROSITE: PS00187; TPP_ENZYMES; 1.
DR PFAM: PF00205; TPP_enzymes; 1.
KW HYPOTHETICAL PROTEIN; LYASE; DECARBOXYLASE; FLAVOPROTEIN;
KW THIAMINE PYROPHOSPHATE.
FT ACT_SITE 49 49 BY SIMILARITY.
SQ SEQUENCE 560 AA; 61288 MW; CEFF14F5 CRC32;
Query Match 71.2%; Score 47; DB 1; Length 560;
Best Local Similarity 57.1%; Pred. No. 8.02e+00;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 333 RODSCWK 339
Qy 116 RHPSCWN 122
Search completed: Thu Jul 8 18:19:45 1999
Job time : 10 secs.

```

THIS PAGE BLANK (USPTO)

W P R E L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:20:03 1999; Maspar time 7.85 Seconds
Tabular output not generated. 48.644 Million cell updates/sec

Title: >US-09-041-236-2
Description: (116-122) from US09041236.pep (8 of 45)
Perfect Score: 66
Sequence: 1 RHPSCWN 7

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 21.909; Variance 27.419; scale 0.799

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	66	100.0	666	4	075326 SEMAPHORIN L.	9.84e-04
2	56	84.8	393	11	088371 SEMAPHORIN L (FRAGMENT	2.09e-01
3	53	80.3	343	10	004980 CYTOCHROME P450 (EC 1.	9.52e-01
4	52	78.8	409	2	031017 ARCA (EC 3.5.3.6) (ARG	1.56e+00
5	51	77.3	187	2	045217 PROTOCATECHUATE-3,4-DI	2.55e+00
6	51	77.3	402	10	023560 HYPOTHETICAL 45.5 KD P	2.55e+00
7	50	75.8	145	1	029747 HYPOTHETICAL 16.7 KD P	4.14e+00
8	49	74.2	77	14	091686 ENVELOPE GLYCOPROTEIN	6.68e+00
9	49	74.2	97	2	087630 HYPOTHETICAL 10.8 KD P	6.68e+00
10	49	74.2	262	14	092434 ACMNPV ORF69.	6.68e+00
11	49	74.2	888	10	004206 DNA-DIRECTED RNA POLYM	6.68e+00
12	48	72.7	77	13	073873 PHOSPHOLIPASE A2 (EC 3	1.07e+01
13	48	72.7	149	13	057313 PHOSPHOLIPASE A2 PRECU	1.07e+01
14	48	72.7	228	2	066125 TRANSPOSASE.	1.07e+01
15	48	72.7	341	13	Q90644 C-TYPE LECTIN (C-TYPE	1.07e+01
16	48	72.7	515	14	090272 GAG PROTEIN.	1.07e+01
17	48	72.7	897	4	060528 ANTIGEN NY-CO-9 (FRAGM	1.07e+01
18	48	72.7	1621	5	001900 CODED FOR BY C. ELEGAN	1.07e+01
19	47	71.2	149	2	068499 TYPE IV PILIN.	1.71e+01
20	47	71.2	329	2	P70784 TECR.	1.71e+01

21	47	71.2	351	4	043515 MUTANT ERYTHROCYTE MEM	1.71e+01
22	47	71.2	367	14	065031 SIGMA B PROTEIN.	1.71e+01
23	47	71.2	367	14	012884 SIGMA B PROTEIN.	1.71e+01
24	47	71.2	413	13	042491 SIMILAR TO HUMAN CDNA	1.71e+01
25	47	71.2	415	4	092552 MYELOBLAST KIAA0264 (F	1.71e+01
26	47	71.2	434	2	P75037 SIMILAR TO GENBANK ACC	1.71e+01
27	47	71.2	534	3	003764 HYPOTHETICAL 61.7 KD P	1.71e+01
28	47	71.2	609	10	Q39477 RNA POLYMERASE SIGMA F	1.71e+01
29	47	71.2	855	10	048544 PHOSPHOLIPASE D-GAMMA	1.71e+01
30	47	71.2	1910	5	Q22426 PROLINE-RICH.	1.71e+01
31	46	69.7	71	8	Q21296 RIBOSOMAL PROTEIN L31.	2.70e+01
32	46	69.7	169	11	070518 HYPOTHETICAL 19.2 KD P	2.70e+01
33	46	69.7	169	11	070517 HYPOTHETICAL 19.2 KD P	2.70e+01
34	46	69.7	171	11	070816 HYPOTHETICAL 19.3 KD P	2.70e+01
35	46	69.7	171	11	070819 HYPOTHETICAL 19.3 KD P	2.70e+01
36	46	69.7	225	14	P89673 CAPSID PROTEIN (FRAGME	2.70e+01
37	46	69.7	256	8	O80029 NADH DEHYDROGENASE SUB	2.70e+01
38	46	69.7	268	14	O56306 ORF69 HOMOLOG.	2.70e+01
39	46	69.7	298	14	O91945 ATTACHMENT PROTEIN.	2.70e+01
40	46	69.7	360	2	O87850 PUTATIVE SECRETED PROT	2.70e+01
41	46	69.7	367	14	O72463 MAJOR OUTER CAPSID PRO	2.70e+01
42	46	69.7	435	4	O75801 BLU PROTEIN TESTIS ISO	2.70e+01
43	46	69.7	465	4	O60260 PARKIN.	2.70e+01
44	46	69.7	591	2	O54204 GENES STRU, STRX, STRV	2.70e+01
45	46	69.7	2127	14	O57294 L PROTEIN, RNA DEPENDE	2.70e+01

ALIGNMENTS

RESULT 1					
ID	075326	PRELIMINARY;	PRT;	566	AA.
AC	075326;				
DT	01-NOV-1998	(TREMREL. 08, CREATED)			
DT	01-NOV-1998	(TREMREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMREL. 08, LAST ANNOTATION UPDATE)			
GN	SEMAPHORIN L.				
GN	SEMAPHORIN L.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;				
OC	CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98389619.				
RA	LANGHE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;				
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA				
RT	viruses."				
RL	GENOMICS 51:340-350(1998).				
DR	EMBL; AF030698; G3523115; --				
SQ	SEQUENCE 566 AA; 74823 MW; F546B39D CRC32;				

Query Match 100.0%; Score 66; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 9.84e-04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	148	RHPSCWN	154
QY	116	RHPSCWN	122

RESULT 2
ID 088371
AC 088371;
PRT; 393 AA.

DT	01-NOV-1998	(TREMREL. 08, CREATED)
DT	01-NOV-1998	(TREMREL. 08, LAST SEQUENCE UPDATE)
DT	01-NOV-1998	(TREMREL. 08, LAST ANNOTATION UPDATE)
DE	SEMAPHORIN L (FRAGMENT).	
GN	SEMAPHORIN L (FRAGMENT).	
OS	MUS MUSCULUS (MOUSE).	
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;	
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 98389619.	

RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.:
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses.";
RL GENOMICS 51:340-350(1998).
DR EMBL: AF030699; G3523117; -.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;

Query Match 84.8%; Score 56; DB 11; Length 393;
Best Local Similarity 85.7%; Pred. No. 2.09e-01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 145 RKPSCWN 151
QY 116 RHPSCWN 122

RESULT 3
ID O04980 PRELIMINARY; PRT; 343 AA.
AC O04980;
DT 01-JUL-1997 (TREMREL. 04, CREATED)
DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 (EC 1.14.14.1) (FRAGMENT).
GN CYP11C1.
OS ZEA MAYS (MAIZE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; ZEA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ETIOLATED COLEOPTILES;
RA SISCO P.H.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC 1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF004210; G2195566; -.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR PFAM: PF00067; p450; 1.
DR MENDEL: 15208; ZEAma.1113; mn15208.
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME.
FT NON_TER 1 1
FT BINDING 283 283 HEME (BY SIMILARITY).
SQ SEQUENCE 343 AA; 38565 MW; 50DB3D3A CRC32;

Query Match 80.3%; Score 53; DB 10; Length 343;
Best Local Similarity 57.1%; Pred. No. 9.52e-01;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 239 RDPSCWN 245
QY 116 RHPSCWN 122

RESULT 4
ID O31017 PRELIMINARY; PRT; 409 AA.
AC O31017;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE ARCA (EC 3.5.3.6) (ARGININE DEIMINASE) (ARGININE DIHYDROLASE).
GN ARCA.
OS RHIZOBIUM ETLI.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; RHIZOBIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98053854.
RA D'HOOGHE I., VANDER WAUWE C., MICHELS J., TRICOT C., DE WILDE P.,
RA VANDERLEYDEN J., STALON V.;
RT "The arginine deiminase pathway in Rhizobium etli: DNA sequence
RT analysis and functional study of the arcABC genes";
RL J. BACTERIOL. 179:7403-7409(1997).
CC 1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O = L-CITRULLINE + NH(3).

DR EMBL: AF025543; G2555170; -.
KW HYDROLASE.
SQ SEQUENCE 409 AA; 45656 MW; CF9D7974 CRC32;

Query Match 78.8%; Score 52; DB 2; Length 409;
Best Local Similarity 83.3%; Pred. No. 1.56e+00;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 158 RDPSCWN 163
QY 116 RHPSCWN 121

RESULT 5
ID Q45217 PRELIMINARY; PRT; 187 AA.
AC Q45217;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE PROTECHUATE-3,4-DIOXYGENASE ALPHA SUBUNIT.
GN PCDA.
OS BRADYRHIZOBIUM JAPONICUM.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; BRADYRHIZOBIUM GROUP;
OC BRADYRHIZOBIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USD110;
RA KOTAGIRI S., SHANTARAM S., PODILA G.K.;
RL SUBMITTED (JUL-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U12426; G521142; -.
KW DIOXYGENASE.
SQ SEQUENCE 187 AA; 20507 MW; B25AEALF CRC32;

Query Match 77.3%; Score 51; DB 2; Length 187;
Best Local Similarity 57.1%; Pred. No. 2.55e+00;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 76 RKPTCWN 82
QY 116 RHPSCWN 122

RESULT 6
ID Q23560 PRELIMINARY; PRT; 402 AA.
AC Q23560;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 45.5 KD PROTEIN.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98121113.
RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,
RA BERGKAMP R., DIRKSE W., VAN STAVEREN M., STIEKEMA W., DROST L.,
RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFFANELLI P., WEDLER H.,
RA WEDLER E., WAMBUIT R., WEITZENEGGER T., POHL T.M., TERRY N.,
RA GIELEN J., VILLARROEL R., DE CLERCK R., VAN MONTAGU M., LECHARNY A.,
RA AUBORG S., GY I., KREIS M., LAO N., KAVANAGH T., HEMPEL S., KOTTER P.,
RA ENTIAN K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,
RA SILVEY M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,
RA VOULELATOU E., MILLIONI D., HATZOPOULOS P., PIRAVANDI E., OBERMAIER B.,
RA HILBERT H., DUESTERHOFT A., MOORES T., JONES J.D.G., ENEVA T.,
RA PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,
RA DELSENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,
RA SCHUELLER C., CHALWATZIS N.;

RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana";
RL NATURE 391:485-488(1998).
DR EMBL: Z97343; E327043; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 402 AA; 45520 MW; 5E678D3 CRC32;

Query Match 77.3%; Score 51; DB 10; Length 402;
Best Local Similarity 83.3%; Pred.No. 2.55e+00;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 11 RPPSCW 16
QY 116 RHPSCW 121

RESULT 7
ID O29747 PRELIMINARY; PRT; 145 AA.
AC O29747;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 16.7 KD PROTEIN.
GN AF0503.
OS ARCHAEOGLOBUS FULGIDUS.
OC ARCHAEA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE;
OC ARCHAEOGLOBUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., OTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic,
RT sulphate-reducing archaeon Archaeoglobus fulgidus";
RL NATURE 390:364-370(1997).
DR EMBL: AE001069; G2650128; -.
DR TIGR: AF0503; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 145 AA; 16721 MW; 4F5FB17B CRC32;

Query Match 75.8%; Score 50; DB 1; Length 145;
Best Local Similarity 56.7%; Pred.No. 4.14e+00;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 117 HPSCWD 122
QY 117 HPSCWN 122

RESULT 8
ID O91686 PRELIMINARY; PRT; 77 AA.
AC O91686;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 2 (HIV-2).
OC VIRUSES; RETROVIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=287;
RA LEITNER T., NORRGREN H., MARQUINA S., AABY P., MELBYE M.,

RA POULSEN A.-G., LARSEN O., DIAS F., ESCANILLA D., ALBERT J.,
RA NAUCL A. JR.;
RT "Hiv-2 genetic variation and DNA load in asymptomatic carriers and
RT AIDS cases in Guinea-Bissau";
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF023888; G3582548; -.
KW ENVELOPE PROTEIN.
FT NON_TER 1
FT NON_TER 77
SQ SEQUENCE 77 AA; 8921 MW; 94A44ADD CRC32;

Query Match 74.2%; Score 49; DB 14; Length 77;
Best Local Similarity 66.7%; Pred.No. 6.68e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 44 RQASCW 49
QY 116 RHPSCW 121

RESULT 9
ID O87630 PRELIMINARY; PRT; 97 AA.
AC O87630;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 10.8 KD PROTEIN.
OS HERBASPIRILLUM SEROPEDICAE.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; BURKHOLDERIA GROUP;
OC HERBASPIRILLUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z78;
RA KLASSEN G., PEDROSA F.O., SOUZA E.M., RIGO L.U.;
RT "Cloning, sequencing and characterization of nifHX-ORF1-ORF2 operon
RT of Herbaspirillum seropedicae";
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF088132; G3644010; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 97 AA; 10776 MW; C778CAB9 CRC32;

Query Match 74.2%; Score 49; DB 2; Length 97;
Best Local Similarity 66.7%; Pred.No. 6.68e+00;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 12 RRPRCW 17
QY 116 RHPSCW 121

RESULT 10
ID O92434 PRELIMINARY; PRT; 262 AA.
AC O92434;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE ACMPV ORF69.
OS BOMBYX MORI NUCLEAR POLYHEDROSIS VIRUS (BMNPV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
OC NUCLEOPOLYHEDROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T3;
RX MEDLINE: 97329351.
RA KAMITA S.G., MAEDA S.;
RT "Sequencing of the putative DNA helicase-encoding gene of the Bombyx
RT mori nuclear polyhedrosis virus and fine-mapping of a region involved
RT in host range expansion";
RL GENE 190:173-179(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=T3;
RA GOMI S., MAJIMA K., MAEDA S.;

"Sequence analysis of the genome of Bombyx mori
RT nucleopolyhedrovirus."
RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: L33180; G3745895; -
SQ SEQUENCE 262 AA; 30384 MW; 419C8C26 CRC32;

Query Match 74.2%; Score 49; DB 14; Length 262;
Best Local Similarity 66.7%; Pred. No. 6.68e+00;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 35 RPRCW 40
QY 116 RHPSCW 121

RESULT 11
ID O04206 PRELIMINARY; PRT; 888 AA.
AC O04206;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE DNA-DIRECTED RNA POLYMERASE II ISOLOG.
GN T29M21.19.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D.; KETCHUM K.A.; LIN X.; PHILLIPS C.A.; BRANDON R.C.;
RA FUHRMANN J.L.; WHITE O.; KERLAVAGE A.R.; ADAMS M.D.; SOMERVILLE C.R.;
RA VENTER J.C.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF002109; G2088656; -
DR PFAM; PF00623; RNA_POL_A; 1.
DR MENDEL; 16065; AKATH; 2935; mn16065.
KW DNA-DIRECTED RNA POLYMERASE.
SQ SEQUENCE 888 AA; 99264 MW; CB1A5A23 CRC32;

Query Match 74.2%; Score 49; DB 10; Length 888;
Best Local Similarity 66.7%; Pred. No. 6.68e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 104 QPGCW 109
QY 117 HPSCW 122

RESULT 12
ID O73873 PRELIMINARY; PRT; 77 AA.
AC O73873;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PHOSPHOLIPASE A2 (EC 3.1.1.4) (FRAGMENT).
OS DICENTRARCHUS LABRAX (EUROPEAN SEA BASS).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;
OC PERCOIDEI; MORONIDAE; DICENTRARCHUS.
RN [1]
RP SEQUENCE FROM N.A.
RN ZAMBONINO INFANTE J.L.; ROLLAND S.E.; CAHU C.L.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AJ006339; E1294705; -
KW HYDROLASE.
FT NON_TER 1
FT NON_TER 77
SQ SEQUENCE 77 AA; 8704 MW; B82EFA2E CRC32;

Query Match 72.7%; Score 48; DB 13; Length 77;
Best Local Similarity 66.7%; Pred. No. 1.07e+01;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 33 QHPECW 38
QY 116 RHPSCW 121

RESULT 13
ID O57313 PRELIMINARY; PRT; 149 AA.
AC O57313;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PHOSPHOLIPASE A2 PRECURSOR.
GN PPLA2-2.
OS CHRYSOPHRYS MAJOR (RED SEA BREAM).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;
OC PERCOIDEI; SPARIDAE; CHRYSOPHRYS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-HEPATOPANCREAS;
RA IJIMA N.; NASU Y.; TAKASHIMA Y.; ESAKA M.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AB009286; D1024642; -
DR PROSITE; PS00118; PA2_HIS; 1.
DR PROSITE; PS00119; PA2_ASP; 1.
KW SIGNAL.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 149 POTENTIAL.
SQ SEQUENCE 149 AA; 16905 MW; AFB3A182 CRC32;

Query Match 72.7%; Score 48; DB 13; Length 149;
Best Local Similarity 66.7%; Pred. No. 1.07e+01;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 80 QHPECW 85
QY 116 RHPSCW 121

RESULT 14
ID O66125 PRELIMINARY; PRT; 228 AA.
AC O66125;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE TRANSPOSASE.
OS BRADYRHIZOBIUM JAPONICUM.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; BRADYRHIZOBIUM GROUP;
OC BRADYRHIZOBIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NK5;
RA ISAWA T.; ITAKURA M.; MINAMISAWA K.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AB003134; D1026192; -
DR PROSITE; PS01007; TRANSPOSASE_MUTATOR; 1.
SQ SEQUENCE 228 AA; 25760 MW; 2D709DD7 CRC32;

Query Match 72.7%; Score 48; DB 2; Length 228;
Best Local Similarity 66.7%; Pred. No. 1.07e+01;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 68 RHQRCW 73
QY 116 RHPSCW 121

RESULT 15
ID Q90644 PRELIMINARY; PRT; 341 AA.
AC Q90644; Q90645;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE C-TYPE LECTIN (C-TYPE LECTIN SHORT FORM).
GN CHBL.
OS GALLUS GALLUS (CHICKEN).
OC EURKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BURSA OF FABRICIUS;
RX MEDLINE; 97461428.
RA GOITSUKA R., CHEN C.H., COOPER M.D.;
RT "B cells in the bursa of Fabricius express a novel C-type lectin
gene.";
RL J. IMMUNOL. 159:3126-3132(1997).
DR EMBL; U22014; G1142650; -;
DR EMBL; U22014; G1142651; -;
KW LECTIN.
SQ SEQUENCE 341 AA; 38311 MW; 676A4D1B CRC32;
Query Match 72.7%; Score 48; DB 13; Length 341;
Best Local Similarity 62.5%; Pred. No. 1.07e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 1; Gaps 1;
Db 57 RHSPGCWS 64
Oy 116 RH-PSCWN 122

Search completed: Thu Jul 8 18:20:16 1999
Job time : 13 secs.

THIS PAGE BLANK (USPTO)

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:24:11 1999; MasPar time 10.62 Seconds
18.025 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (120-128) from US09041236.ppt (9 of 45)
Perfect Score: 75
Sequence: 1 CWNLVNGTV 9

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 17.327; Variance 57.862; scale 0.299

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	56	74.7	638	5	E. coli HSP (dnaK)	4.88e+01
2	55	73.3	265	28	Amino acid sequence o	6.13e+01
3	55	73.3	265	17	Mature inositol-monop	6.13e+01
4	55	73.3	288	35	IMP-18p myo-inositol	6.13e+01
5	55	73.3	339	39	Kidney injury associa	6.13e+01
6	54	72.0	709	4	Toxic KHS gene.	7.70e+01
7	52	69.3	323	21	Yeast transcription r	1.21e+02
8	50	66.7	292	29	Methanococcus thermol	1.89e+02
9	49	65.3	162	10	Human GR C-terminal (2.36e+02
10	49	65.3	477	10	Human glucagon recept	2.36e+02
11	49	65.3	485	10	Rat glucagon receptor	2.36e+02
12	49	65.3	541	21	G-protein parathyroid	2.36e+02
13	48	64.0	20	29	Dopamine transporter	2.94e+02
14	48	64.0	23	5	PTH/PTHrP receptor fr	2.94e+02
15	48	64.0	26	38	Parathyroid hormone r	2.94e+02
16	48	64.0	26	16	PTH/PTHrP extracellul	2.94e+02

17	48	64.0	74	31	W28173	Staphylococcus aureus	2.94e+02
18	48	64.0	107	34	W77670	Staphylococcus aureus	2.94e+02
19	48	64.0	506	39	W87626	Dirofilaria immitis a	2.94e+02
20	48	64.0	515	38	W73314	Parathyroid hormone r	2.94e+02
21	48	64.0	515	16	R92275	Opoussum kidney PTH/PT	2.94e+02
22	48	64.0	515	5	R27704	Opoussum kidney PTH/PT	2.94e+02
23	48	64.0	517	26	W35713	Japanese morning glori	2.94e+02
24	48	64.0	585	16	R92276	Opoussum kidney PTH/PT	2.94e+02
25	48	64.0	585	38	W73315	Parathyroid hormone r	2.94e+02
26	48	64.0	585	5	R27705	Opoussum kidney PTH/PT	2.94e+02
27	48	64.0	590	39	W87628	Dirofilaria immitis a	2.94e+02
28	48	64.0	590	39	W87629	Dirofilaria immitis a	2.94e+02
29	48	64.0	591	38	W73316	Parathyroid hormone r	2.94e+02
30	48	64.0	591	5	R27706	Rat bone PTH/PTHrP re	2.94e+02
31	48	64.0	591	16	R92277	Rat bone PTH/PTHrP re	2.94e+02
32	48	64.0	593	38	W73317	Human Parathyroid hor	2.94e+02
33	48	64.0	593	16	R92278	Human kidney PTH/PTHr	2.94e+02
34	48	64.0	614	5	R27707	Human kidney PTH/PTHr	2.94e+02
35	48	64.0	619	7	R36395	Dopamine transporter	2.94e+02
36	48	64.0	620	8	R44492	Predicted sequence of	2.94e+02
37	48	64.0	1027	22	W17057	Candida albicans chit	2.94e+02
38	48	64.0	1422	16	R82071	Hepatitis GB virus (H	2.94e+02
39	47	62.7	164	38	W82542	Human IL-1 delta prot	3.66e+02
40	47	62.7	245	16	R81416	Hepatitis GB virus (H	3.66e+02
41	47	62.7	586	23	W10423	Cytomegalovirus U84	3.66e+02
42	47	62.7	763	28	W31852	Mycobacterium tubercu	3.66e+02
43	47	62.7	797	31	W56380	Human GlyT-2 transpor	3.66e+02
44	47	62.7	797	31	W56386	Human GlyT-2 transpor	3.66e+02
45	47	62.7	797	31	W56383	Human GlyT-2 transpor	3.66e+02

ALIGNMENTS

RESULT 1

ID R03924 standard; Protein; 638 AA.
AC R03924;
DT 30-AUG-1990 (first entry)
DE E. coli HSP (dnaK).
KW Hsp70; heat shock protein; Trypanosoma cruzi; vaccines; HSP.
OS Escherichia coli.
PN W09002564-A.
PD 22-MAR-1990.
PF 12-SEP-1989; 03955.
PR 12-SEP-1988; US-243474.
PA (CODO-) Codon.
PI Dragon E, Faulds D, Sias S;
DR WPI; 90-115820/15.
PT Proteins homologous to heat shock proteins from Trypanosoma cruzi - used
PT in vaccines and diagnosis for species of eg Mycoplasma or Mycobacteria.
PS Disclosure; Fig 2.1.2.14; 86pp; English.
CC Fig. 2 provides an alignment of heat shock proteins from a variety
CC of organisms: 1. M.hypopneumoniae (Mhysp70 - R03922);
CC 2. Bacillus megaterium (Bmebsp70 - R03923);
CC 3. E. coli (dnaK - R03924);
CC 4. T. cruzi (tc70kd - R03925);
CC 5. T. cruzi (R03926);
CC 6. Rat rattus (rathsp70 - R03927);
CC 7. Xenopus laevis (xl70 - R03928);
CC 8. Homo sapiens (humbsp70 - R03929);
CC 9. Gallus gallus (chkhsp70 - R03930);
CC 10. Zea mays (zmehsp70 - R03931);
CC 11. Serratia marcescens (smahsp70 - R03932).
CC The proteins having homology to hsp's of T. cruzi can be used in
CC vaccines and diagnosis involving e.g. Trypanosoma, Mycoplasma and
CC Mycobacteria species.
SQ Sequence 638 AA;

Query Match 74.7%; Score 56; DB 5; Length 638;
Best Local Similarity 55.6%; Pred. No. 4.88e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 309 cowlvnrsi 317
|||||:

```

Qy 120 CWNLVNGTV 128

RESULT 2
ID W46518 standard; Protein; 265 AA.
AC W46518;
DT 20-MAY-1998 (first entry)
DE Amino acid sequence of human inositol monophosphatase-H1 cDNA.
KW inositol monophosphatase-H1; hIMP-H1; inositol phosphatase;
KW inositol monophosphatase; screening; inhibition; enzyme; treatment;
KW psychiatric disorder; antibody; antagonist; psychotic disorder;
KW depressive disorder.
KW Homo sapiens.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 168 /note= "encoded by GAG"
FT
PN US5716806-A.
PD 10-FEB-1998. 461731.
PF 05-JUN-1995; 461731.
PR 05-JUN-1995; US-461731.
PR 16-SEP-1994; WO-U10465.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Gocayne JD, Meissner PS;
DR WPI: 98-158367/14.
DR N-PSDB: V05296.
PT Nucleic acid encoding human inositol mono-phosphatase H1 polypeptide
PT - useful for producing recombinant polypeptide
PS Claim 1: Fig 1: 23pp: English.
CC The present sequence represents a human inositol monophosphatase-H1
CC (hIMP-H1) protein. The nucleotide sequence was isolated from a cDNA
CC library derived from human brain tissue. The hIMP-H1 protein is
CC structurally related to the inositol phosphatase family, and exhibits
CC the highest degree of homology to human inositol monophosphatase.
CC The cDNA sequence is cloned to produce recombinant vectors which
CC express a mature polypeptide, as well as its biologically active and
CC diagnostically or therapeutically useful fragments, analogues and
CC derivatives. The polypeptides or polynucleotides are used screening
CC and designing compounds capable of inhibiting a class of enzymes, and
CC the treatment of psychiatric disorders. The protein can be used to raise
CC antibodies or antagonists, which may be used to inhibit the action of
CC such proteins e.g. in the treatment of psychotic and depressive
CC disorders (bipolar and non-bipolar).
SQ Sequence 265 AA;

Query Match 73.3%; Score 55; DB 28; Length 265;
Best Local Similarity 55.6%; Pred. No. 6.13e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 206 cwdlaaatv 214
||| :||
Qy 120 CWNLVNGTV 128

RESULT 3
ID R93156 standard; Protein; 265 AA.
AC R93156;
DT 17-AUG-1996 (first entry)
DE Mature inositol-monophosphatase-H1.
KW Human; inositol-monophosphatase-H1; brain; drug screening;
KW psychosis; manic depression; antipsychotic; antidepressant;
KW mapping; gene therapy; triple helix; antisense; antibody;
KW diagnostic.
KW Homo sapiens.
OS Homo sapiens.
FH Key Location/Qualifiers
FT active_site 78..83
FT /note= "Conserved residues essential at inositol-
FT monophosphatase active sites"
FT
PN WO9608557-A1.
PD 21-MAR-1996.
PF 16-SEP-1994; WO-U10465.
PR 16-SEP-1994; WO-U10465.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Gocayne JD, Meissner PS;

WPI: 96-179934/18.
DR N-PSDB: T17178, T17179.
PT New isolated human inositol mono:phosphatase H1 and DNA - used to
PT develop prods. for the diagnosis and therapy of psychiatric
PT disorders, e.g. psychotic and depressive disorders
PS Claim 1; Page 35-36; 51pp: English.
CC The sequence represents mature human inositol-monophosphatase-H1,
CC encoded by a gene isolated from a human brain cDNA library. The
CC sequence contains a conserved sequence essential at the active site
CC of inositol-monophosphatases. The enzyme plays an important role
CC in the phosphatidylinositol signalling pathway by catalysing the
CC hydrolysis of inositol monophosphates, and may be the molecular
CC site of action for lithium therapy for manic depressive illness.
CC The enzyme may be used to screen drugs for inhibitor activity, for
CC use in therapy of psychiatric disorders, e.g. psychosis or manic
CC depression. The DNA may be used in mapping or gene therapy (e.g.
CC via triple helix or antisense strategies) of genetic disease.
CC Antibodies against the enzyme may be used as diagnostic or
CC therapeutic agents.
SQ Sequence 265 AA;

Query Match 73.3%; Score 55; DB 17; Length 265;
Best Local Similarity 55.6%; Pred. No. 6.13e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 206 cwdlaaatv 214
||| :||
Qy 120 CWNLVNGTV 128

RESULT 4
ID W60676 standard; Protein; 288 AA.
AC W60676;
DT 18-NOV-1998 (first entry)
DE IMP-18p myo-inositol monophosphatase protein.
KW Manic-depressive illness; susceptibility; genotype; diagnosis;
KW chromosomal marker; polymorphic marker; chromosome 18; human;
KW myo-inositol monophosphatase protein; IMP-18p.
OS Homo sapiens.
PN WO9818963-A1.
PD 07-MAY-1998.
PF 28-OCT-1997; U19381.
PR 28-OCT-1996; US-029278.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Badner JA, Berrettini WH, Detera-Wadleigh SD, Esterling LE,
PI Garshon ES, Goldin LR, Sanders AR, Yoshikawa T;
DR WPI: 98-272247/24.
DR N-PSDB: V38343.
PT New isolated IMP-18p myo-inositol monophosphatase - used to develop
PT products for determining susceptibility to manic depressive illness
PT and as targets for preventive and therapeutic treatments
PS Claim 20; Fig 5; 118pp: English.
CC A method has been developed for determining a genotype associated with
CC increased susceptibility to manic-depressive (MD) illness. The method
CC comprises determining the genotype of an affected individual with at
CC least one polymorphic marker localised within the chromosomal region
CC defined by and including markers D18S843 and D18S869 and determining the
CC genotype associated with increased susceptibility to MD disorder. The
CC method can be used for determining susceptibility to MD illness
CC including bipolar disorder, genetic counselling of individuals from
CC families affected with MD illness, and aid in the differential diagnosis
CC of MD illness from other psychiatric pathologies. Products from the
CC present invention can also be used to obtain modulators of IMP-18p myo-
CC inositol monophosphatase protein activity and as targets for preventive
CC and therapeutic treatments. The present sequence represents the
CC specifically claimed IMP-18p myo-inositol monophosphatase protein
CC isolated from chromosome 18.
SQ Sequence 288 AA;

Query Match 73.3%; Score 55; DB 35; Length 288;
Best Local Similarity 55.6%; Pred. No. 6.13e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

Db 229 cwdlaaatv 237
 ||:| :||
 QY 120 CWNLVNGTV 128

RESULT 5

ID W86325 standard; Protein: 339 AA.
 AC W86325;
 DT 01-MAR-1999 (first entry)
 DE Kidney injury associated molecule HW057 protein.
 KW Kidney injury associated molecule; kidney injury related molecule;
 KW KIM; tissue growth promotion; regeneration; renal condition;
 KW acute renal failure; acute nephritis; tumour.
 OS Rattus sp.
 PN W09853071-A1.
 PD 26-NOV-1998.
 PF 22-MAY-1998: U10547.
 PR 23-MAY-1997: US-047491.
 PR 23-MAY-1997: US-047490.
 PA (BIOJ) BIOGEN INC.
 PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
 DR WPI: 99-045312/04.
 DR N-PSDB; V80610.
 PT Kidney injury-associated molecule, KIM, polypeptides - upregulated
 PT in injured or regenerating tissues, useful to promote tissue growth
 PT and regeneration, especially to treat renal conditions
 PS Claim 17; Page 131-132; 213pp; English.
 CC The present sequence represents a kidney injury associated molecule
 CC (KIM) protein. KIM proteins can be administered therapeutically
 CC by expressing KIM encoding polynucleotides, to promote growth and/or
 CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins
 CC are upregulated in injured or regenerating (especially renal) tissues.
 CC KIM fusion proteins, conjugates, antibodies and vectors can also be used
 CC therapeutically, e.g. these or the KIM proteins may be included with an
 CC acceptable carrier in pharmaceutical compositions, useful for therapy/
 CC prophylaxis of conditions associated with dysfunction/disregulation of
 CC KIM genes or proteins, especially renal diseases or impairments of renal
 CC function in humans (e.g. acute renal failure, acute nephritis). The
 CC polynucleotides can be used to produce antisense sequences which, when
 CC internalised into cells, can disrupt expression of a cellular KIM gene,
 CC also useful in therapy (e.g. to block the growth of tumours dependent on
 CC KIM for growth) or compositions. The proteins and polynucleotides are
 CC useful diagnostically e.g. to detect and quantify renal injury/disease
 CC (indicative of increased risk, or presence of, renal injury or impaired
 CC function), or abnormal responses to tissue injury (indicative of
 CC increased risk, or presence of, an autoimmune response or abnormal
 CC tissue growth arising from/affecting renal tissue). The proteins can
 CC also be used to locate KIM-producing cells (especially specific loci,
 CC e.g. tissue masses abnormally producing/expressing KIM such as tumours
 CC arising from/affecting renal tissue), by contacting cells with an
 CC imageable KIM-binding reagent and imaging reagent accumulation.
 SQ Sequence 339 AA;

Query Match 73.3%; Score 55; DB 39; Length 339;
 Best Local Similarity 55.6%; Pred. No. 6.13e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 280 cwdlaaatv 288
 ||:| :||
 QY 120 CWNLVNGTV 128

RESULT 6

ID R24130 standard; Protein: 709 AA.
 AC R24130;
 DT 14-NOV-1992 (first entry)
 DE Toxic KHS gene.
 KW Toxic KHS; autoimmunity; Candida glabrata.
 OS Saccharomyces cerevisiae.
 PN J04112792-A.
 PD 14-APR-1992.
 PF 03-SEP-1990: 230605.
 PR 03-SEP-1990; JP-230605.

PA (CHOK/) CHOKAN K.
 DR WPI: 92-173148/21.
 DR N-PSDB; Q25052.
 PT cDNA sequence contg. promoter region and coding region of a toxic KHS
 PT gene - for prepn. of KHS toxin in large amts.
 PS Disclosure; Fig 3; 8pp; Japanese.
 CC The sequence given is derived from a DNA sequence containing the
 CC promoter region and the coding region of a toxic KHS gene. This
 CC DNA sequence can be introduced into a bacterial host, eg E. coli,
 CC with transformants being identified by the autoimmunity contained in
 CC the DNA sequence. The selected transformants were found to inhibit
 CC the growth of Candida glabrata.
 SQ Sequence 709 AA;

Query Match 72.0%; Score 54; DB 4; Length 709;
 Best Local Similarity 77.8%; Pred. No. 7.70e+01;
 Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 699 cw-lyngv 706
 ||:| :||
 QY 120 CWNLVNGTV 128

RESULT 7

ID W13828 standard; Protein: 323 AA.
 AC W13828;
 DT 04-JUN-1997 (first entry)
 DE Yeast transcription regulatory factor SRB11.
 KW Transcription regulatory factor; suppressor of RNA polymerase B;
 KW SRB11; RNA polymerase II; holoenzyme; SWI/SNF.
 OS Saccharomyces cerevisiae.
 PN W09708301-A1.
 PD 06-MAR-1997.
 PF 28-AUG-1996; U14192.
 PR 31-AUG-1995; US-521872.
 PR 11-OCT-1995; US-540804.
 PR 26-JAN-1996; US-590399.
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PI Chao DM, Koleske AJ, Thompson CM, Young RA;
 DR WPI: 97-179258/16.
 DR N-PSDB; T59911.

PT Purified RNA polymerase II holoenzyme - comprises RNA polymerase II
 PT and one or more regulatory proteins, pref. suppressor of RNA
 PT polymerase B proteins or SWI/SNF proteins
 PS Claim 11; Fig 13; 154pp; English.
 CC Novel yeast SRB (suppressor of RNA polymerase B) proteins SRB4, SRB5,
 CC SRB6, SRB7, SRB8, SRB9, SRB10 and SRB11 (W13821-28) are transcriptions
 CC regulatory factors that act as positive and negative regulators of
 CC RNA polymerase II activity, and are components of the RNA polymerase
 CC II holoenzyme. They were identified using methods designed to
 CC identify transcription factors involved in RNA polymerase II
 CC C-terminal domain (CTD) function. Genomic clones (T59904-11) for
 CC the SRBs have been obtd. SRBs can be used to treat diseases
 CC resulting from alteration or deletion of the SRB gene, pref. by gene
 CC transfer technology. They can also be used in vitro
 CC transcription of DNA and to identify cpds. that modify gene
 CC transcription.
 SQ Sequence 323 AA;

Query Match 69.3%; Score 52; DB 21; Length 323;
 Best Local Similarity 50.0%; Pred. No. 1.21e+02;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 208 cwsllnds 215
 ||:| :||
 QY 120 CWNLVNGTV 127

RESULT 8

ID W42389 standard; Protein: 292 AA.
 AC W42389;
 DT 22-JUN-1998 (first entry)
 DE Methanococcus thermoautotrophicus phosphatase (14phl).

KW Alkaline phosphatase; thermostable enzyme; thermophilic bacterium;
KW food; detergent; baking.
OS Methanococcus thermolithoautotrophicus strain SN1.
PN WO9748416-A1.
PD 24-DEC-1997.
PF 19-JUN-1997; U10784
PR 19-JUN-1996; US-033752.
PI (RECO-) RECOMBINANT BIOCATALYSIS INC.
PI Bylina E, Lee E, Mathur EJ;
DR WPI: 98-062851/06.
DR N-PSDB: V03314.
PT Thermostable phosphatase(s) - useful in pharmaceutical, food,
PT detergent, and baking industries
PS Claim 11; Page 89-90; 128pp; English.
CC This protein comprises a thermostable phosphatase, designated
CC 14ph1, of Methanococcus thermolithoautotrophicus. The invention
CC relates to claimed polynucleotides (see V03301-20) coding for
CC claimed thermostable phosphatases (see W42380-95). Vector and host
CC cells are used to produce the enzymes, which can be used in a
CC claimed method to hydrolyse phosphate bonds. They can also be used
CC in enzyme labelling processes, in certain recombinant DNA
CC techniques, in ELISA immunoassays, in enzyme linked gene probes,
CC in research applications for removing 5' phosphates in
CC polynucleotides prior to end labelling, and in the pharmaceutical,
CC food, detergent, and baking industries.
SQ Sequence 292 AA;

Query Match 66.7%; Score 50; DB 29; Length 292;
Best Local Similarity 50.0%; Pred. No. 1.89e+02; Indels 0; Gaps 0;
Matches 4; Conservative 3; Mismatches 1;

Db 234 wdlldgvd 241
I:::I
QY 121 WNLVNGTV 128
I:::I

RESULT 9
ID R57829 standard; Protein; 162 AA.
AC R57829;
DT 03-OCT-1994 (first entry)
DE Human GR C-terminal (clone 40-2-2).
KW Rat; human; glucagon receptor; transgenic animal; metabolism; model;
KW amplification; primer; polymerase chain reaction; PCR.
OS Homo sapiens.
PN WO9405789-A.
PD 17-MAR-1994.
PF 30-AUG-1993; U08174.
PR 28-AUG-1992; US-938331.
PR 01-JUL-1993; US-086631.
PA (ZYMO) ZYMOGENETICS INC.
PI Foster DC, Grant FJ, Jelinek LJ, Kindsvogle WR;
PI Kuijper JL, Lok S, O'Hara PJ, Sheppard PO;
DR WPI: 94-101194/12.
DR N-PSDB: Q67247.
PT New recombinant glucagon receptors and antibodies - useful to
PT produce model transgenic animals for study and with therapeutic
PT applications
PS Example 5; Page 86-87; 112pp; English.
CC Example 5 describes the isolation of human GR.
CC The GR coding sequence was constructed from partial clone
CC p9A11 (5' end) and clone 40-2-2 (3' end).
CC Host cells contg. GR DNA may be used for the prodn. of
CC recombinant GR. GR DNA may also be expressed in non-human
CC transgenic animals, pref. mice. Such animals may be readily
CC used as models to study the role of the glucagon receptor
CC in metabolism.
SQ Sequence 162 AA;

Query Match 65.3%; Score 49; DB 10; Length 162;
Best Local Similarity 71.4%; Pred. No. 2.36e+02; Indels 0; Gaps 0;
Matches 5; Conservative 1; Mismatches 1;

Db 96 cwliveg 102

KW Alkaline phosphatase; thermostable enzyme; thermophilic bacterium;
KW food; detergent; baking.
OS Methanococcus thermolithoautotrophicus strain SN1.
PN WO9748416-A1.
PD 24-DEC-1997.
PF 19-JUN-1997; U10784
PR 19-JUN-1996; US-033752.
PI (RECO-) RECOMBINANT BIOCATALYSIS INC.
PI Bylina E, Lee E, Mathur EJ;
DR WPI: 98-062851/06.
DR N-PSDB: V03314.
PT Thermostable phosphatase(s) - useful in pharmaceutical, food,
PT detergent, and baking industries
PS Claim 11; Page 89-90; 128pp; English.
CC This protein comprises a thermostable phosphatase, designated
CC 14ph1, of Methanococcus thermolithoautotrophicus. The invention
CC relates to claimed polynucleotides (see V03301-20) coding for
CC claimed thermostable phosphatases (see W42380-95). Vector and host
CC cells are used to produce the enzymes, which can be used in a
CC claimed method to hydrolyse phosphate bonds. They can also be used
CC in enzyme labelling processes, in certain recombinant DNA
CC techniques, in ELISA immunoassays, in enzyme linked gene probes,
CC in research applications for removing 5' phosphates in
CC polynucleotides prior to end labelling, and in the pharmaceutical,
CC food, detergent, and baking industries.
SQ Sequence 292 AA;

Query Match 66.7%; Score 50; DB 29; Length 292;
Best Local Similarity 50.0%; Pred. No. 1.89e+02; Indels 0; Gaps 0;
Matches 4; Conservative 3; Mismatches 1;

Db 234 wdlldgvd 241
I:::I
QY 121 WNLVNGTV 128
I:::I

RESULT 9
ID R57829 standard; Protein; 162 AA.
AC R57829;
DT 03-OCT-1994 (first entry)
DE Human GR C-terminal (clone 40-2-2).
KW Rat; human; glucagon receptor; transgenic animal; metabolism; model;
KW amplification; primer; polymerase chain reaction; PCR.
OS Homo sapiens.
PN WO9405789-A.
PD 17-MAR-1994.
PF 30-AUG-1993; U08174.
PR 28-AUG-1992; US-938331.
PR 01-JUL-1993; US-086631.
PA (ZYMO) ZYMOGENETICS INC.
PI Foster DC, Grant FJ, Jelinek LJ, Kindsvogle WR;
PI Kuijper JL, Lok S, O'Hara PJ, Sheppard PO;
DR WPI: 94-101194/12.
DR N-PSDB: Q67247.
PT New recombinant glucagon receptors and antibodies - useful to
PT produce model transgenic animals for study and with therapeutic
PT applications
PS Example 5; Page 86-87; 112pp; English.
CC Example 5 describes the isolation of human GR.
CC The GR coding sequence was constructed from partial clone
CC p9A11 (5' end) and clone 40-2-2 (3' end).
CC Host cells contg. GR DNA may be used for the prodn. of
CC recombinant GR. GR DNA may also be expressed in non-human
CC transgenic animals, pref. mice. Such animals may be readily
CC used as models to study the role of the glucagon receptor
CC in metabolism.
SQ Sequence 162 AA;

Query Match 65.3%; Score 49; DB 10; Length 162;
Best Local Similarity 71.4%; Pred. No. 2.36e+02; Indels 0; Gaps 0;
Matches 5; Conservative 1; Mismatches 1;

Db 96 cwliveg 102

QY 120 CWNLVNG 126
I:::I

RESULT 10
ID R50047 standard; Protein; 477 AA.
AC R50047;
DT 03-OCT-1994 (first entry)
DE Human glucagon receptor.
KW Rat; human; glucagon receptor; transgenic animal; metabolism; model;
KW signal sequence; transmembrane domain; glycosylation site.
OS Homo sapiens.
PN WO9405789-A.
PD 17-MAR-1994.
PF 30-AUG-1993; U08174.
PR 28-AUG-1992; US-938331.
PR 01-JUL-1993; US-086631.
PA (ZYMO) ZYMOGENETICS INC.
PI Foster DC, Grant FJ, Jelinek LJ, Kindsvogle WR;
PI Kuijper JL, Lok S, O'Hara PJ, Sheppard PO;
DR WPI: 94-101194/12.
DR N-PSDB: Q58776.
PT New recombinant glucagon receptors and antibodies - useful to
PT produce model transgenic animals for study and with therapeutic
PT applications
PS Claim 4; Page 91-95; 112pp; English.
CC Rat and human glucagon receptor (GR) DNA was isolated.
CC Host cells contg. GR DNA may be used for the prodn. of
CC recombinant GR. GR DNA may also be expressed in non-human
CC transgenic animals, pref. mice. Such animals may be readily
CC used as models to study the role of the glucagon receptor
CC in metabolism.
SQ Sequence 477 AA;

Query Match 65.3%; Score 49; DB 10; Length 477;
Best Local Similarity 71.4%; Pred. No. 2.36e+02; Indels 0; Gaps 0;
Matches 5; Conservative 1; Mismatches 1;

Db 240 cwliveg 246
I:::I
QY 120 CWNLVNG 126
I:::I

RESULT 11
ID R50046 standard; Protein; 485 AA.
AC R50046;
DT 03-OCT-1994 (first entry)
DE Rat glucagon receptor.
KW Rat; human; glucagon receptor; transgenic animal; metabolism; model;
KW signal sequence; transmembrane domain; glycosylation site.
OS Rattus rattus.
PN WO9405789-A.
PD 17-MAR-1994.
PF 30-AUG-1993; U08174.
PR 28-AUG-1992; US-938331.
PR 01-JUL-1993; US-086631.
PA (ZYMO) ZYMOGENETICS INC.
PI Foster DC, Grant FJ, Jelinek LJ, Kindsvogle WR;
PI Kuijper JL, Lok S, O'Hara PJ, Sheppard PO;
DR WPI: 94-101194/12.
DR N-PSDB: Q58758.
PT New recombinant glucagon receptors and antibodies - useful to
PT produce model transgenic animals for study and with therapeutic
PT applications
PS Claim 4; Page 77-81; 112pp; English.
CC Rat and human glucagon receptor (GR) DNA was isolated.
CC Host cells contg. GR DNA may be used for the prodn. of
CC recombinant GR. GR DNA may also be expressed in non-human
CC transgenic animals, pref. mice. Such animals may be readily
CC used as models to study the role of the glucagon receptor
CC in metabolism.
CC The rat receptor comprises 8 clusters of hydrophobic amino
CC acids corresp. to an N-terminal signal sequence and seven

CC transmembrane domains. 4 potential N-linked glycosylation
 CC sites are located in an extended hydrophilic sequence.
 SQ Sequence 485 AA;

Query Match 65.3%; Score 49; DB 10; Length 485;
 Best Local Similarity 71.4%; Pred. No. 2.36e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 241 cwliveg 247
 |||||
 QY 120 CWNLVNG 126

RESULT 12
 ID W12695 standard; Protein; 541 AA.
 AC W12695;
 DT 31-MAY-1997 (first entry)
 DE G-protein parathyroid hormone receptor HLTGDG74.
 KW G-protein parathyroid hormone receptor; HLTGDG74; parathormone; PTH;
 KW calcium; signal transduction; agonist; antagonist; hypocalcaemia;
 KW hyperphosphataemia; hypoparathyroidism; chronic tetany;
 KW osteoporosis; hypercalcaemia; hypoparathyroidism; hypophosphataemia;
 KW kidney stone; nephrolithiasis; therapy; diagnosis.
 OS Homo sapiens.
 PN WO9639433-A1.
 PD 12-DEC-1996.
 PF 05-JUN-1995; U07085.
 PR 05-JUN-1995; WO-U07085.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Li Y, Rosen CA, Ruben SM, Soppet DR;
 DR N-PSDB; T59619.
 PT Human G-protein parathyroid hormone receptor, HLTGDG74 - used to
 PT identify (ant)agonists, used in the treatment of hypo- or
 PT hyper-calcaemia, hypo- or hyper-phosphatemia, kidney stones, etc
 PS Claim 9; Fig 1A-E; 62pp; English.
 CC A novel 7-transmembrane receptor (W12695) has been identified as a
 CC human G-protein parathyroid hormone (PTH) receptor, designated
 CC HLTGDG74. It shows 48.2% homology to the human PTH receptor. Its
 CC amino acid sequence was deduced from a cDNA clone (T59619) isolated
 CC from a human T cell lymphoma tissue cDNA library. Recombinant
 CC HLTGDG74 can be produced in transformed host cells and used to
 CC screen for (ant)agonist cpds. Agonists may be used to prevent or
 CC treat e.g. hypocalcaemia, hyperphosphataemia, hypoparathyroidism
 CC and chronic tetany by stimulating an increase in serum calcium
 CC levels. Antagonists can be used to inhibit the receptor e.g. for
 CC the treatment of osteoporosis, hypercalcaemia, hypoparathyroidism,
 CC hypophosphataemia, kidney stone, nephrolithiasis.
 SQ Sequence 541 AA;

Query Match 65.3%; Score 49; DB 21; Length 541;
 Best Local Similarity 44.4%; Pred. No. 2.36e+02;
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 306 cwlvsagdi 314
 |||||
 QY 120 CWNLVNGTV 128

RESULT 13
 ID W33986 standard; peptide; 20 AA.
 AC W33986;
 DT 28-MAY-1998 (first entry)
 DE Dopamine transporter protein antagonist.
 KW Antagonist; integral membrane protein; inhibitor; transmembrane domain;
 KW therapy; psychotic disorder; Huntington's disease; Parkinson's disease;
 KW post-myocardial infarction; HIV infection; cell proliferation disorder;
 KW peptic ulcer; glaucoma; asthma; renal disease; autoimmune disease;
 KW antibacterial; dopamine transporter protein.
 OS Homo sapiens.
 PN WO9735881-A2.
 PD 02-OCT-1997.
 PF 26-MAR-1997; CA0203.

PR 20-AUG-1996; US-024240.
 PR 27-MAR-1996; US-014306.
 PR 25-JUN-1996; US-670119.
 PA (GEOR/) GEORGE S R.
 PA (NGGY/) NG G Y K.
 PA (ODOW/) O'DOWD B F.
 PA (SEEM/) SEEMAN P.
 PI George SR, Ng GYK, ODOWD BF, Seeman P;
 DR WPI; 97-489566/45.
 PT Inhibitor of integral membrane protein - used to treat e.g.
 PT schizophrenia, hypertension, viral infection, cancer etc.
 PS Claim 30; Page 97; 127pp; English.
 CC This sequence is a dopamine transporter protein antagonist, and is an
 CC antagonist of the invention. The antagonists (A) are for inhibiting the
 CC function of a prokaryotic or eukaryotic integral membrane protein (IMP)
 CC having at least one transmembrane domain (TMD), comprises a peptide
 CC including at least 4 consecutive amino acids (aa) from the sequence of
 CC the TMD. (A) are used to treat or prevent disorders in mammals that
 CC involve disturbances of IMP, and the same effect is achieved by inserting
 CC a nucleic acid encoding (A) into the cells of the mammal. Specified
 CC conditions that can be treated are schizophrenia, psychotic disorders,
 CC Huntington's disease, Tourette's syndrome, hypertension, post-myocardial
 CC infarction, tachyarrhythmia, human immunodeficiency virus infection,
 CC AIDS, cell proliferation (cancer, psoriasis or hyperkeratosis), and drug
 CC abuse. (A) can also be used to treat peptic ulcer, glaucoma, asthma,
 CC renal disease, Parkinson's disease, congestive heart failure, migraine,
 CC depression, obesity, diabetic neuropathy/retinopathy, autoimmune
 CC diseases, graft rejection etc. (A) can be used as diuretics and
 CC antibacterials. (A) may also be used to determine function of
 CC orphan receptors and, when attached to a marker, for tissue imaging to
 CC localise/quantify specific receptors. The transgenic animals are used to
 CC test efficacy/toxicity of (A) and as models of diseases. (A) are very
 CC specific and selective, against the IMP from which they are derived, so
 CC should have few side effects.
 SQ Sequence 20 AA;

Query Match 64.08; Score 48; DB 29; Length 20;
 Best Local Similarity 66.7%; Pred. No. 2.94e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 2 cwlxlv 7
 |||||
 QY 120 CWNLVN 125

RESULT 14
 ID R27711 standard; peptide; 23 AA.
 AC R27711;

DT 16-MAR-1993 (first entry)
 DE PTH/PTHrP receptor fragment.
 KW Parathyroid hormone; related protein; calcium; antagonist;
 KW antibodies; hypercalcaemia; extracellular domain.
 OS Synthetic.
 PN WO9217602-A.
 PD 15-OCT-1992.
 PF 06-APR-1992; U02821.
 PR 05-APR-1991; US-681702.
 PR 06-APR-1992; US-864475.

PA (GEOH) GEN HOSPITAL CORP OFFICE TECHNOLOGY.
 PI Abou-samra A, Juppner H, Kronenberg HM, Potts JT, Schipani E;
 PI Segre GV;
 DR WPI; 92-366271/44.
 PT New DNA encoding parathyroid hormone receptor, DNA and antibodies
 PT - for (differential) diagnosis of hypercalcaemia, and diagnosis
 PT and treatment of tumours
 PS Claim 25; Page 5; 91pp; English.
 CC The peptide sequence shown represents an extracellular fragment of
 CC a parathyroid hormone/parathyroid hormone related protein
 CC (PTH/PTHrP) receptor protein. The peptide is capable of binding
 CC PTH or PTHrP and acting as an antagonist of these cpds. The
 CC peptide may be used to inhibit activation of PTH or PTHrP and thus
 CC reduce the level of calcium in the blood. Cpds. capable of competing
 CC with PTH or PTHrP for binding can be identified using the protein prod.

CC and DNAs homologous to PTH DNA can be identified using fragments of the
CC clone as probes. The sequence may be used for the prodn. of antibodies
CC useful for the treatment, classification, prognosis and/or treatment of
CC disorders related to the interaction between a cell receptor and a
CC ligand such as in hypercalcaemia. See also R27704-16.
SQ Sequence 23 AA;

Query Match 64.0%; Score 48; DB 5; Length 23;
Best Local Similarity 57.1%; Pred. No. 2.94e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 cwdlssg 15
||:|:|
QY 120 CWNLVNG 126

RESULT 15

ID W73321 standard; peptide: 26 AA.
AC W73321;
DT 08-FEB-1999 (first entry)
DE Parathyroid hormone receptor fragment RP-4.
KW Parathyroid hormone receptor; PTH receptor; antibody; therapy;
KW PTH-related hypercalcaemia; opossum; rat; human.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc_difference 2 /note="unspecified amino acid"
PN US5840853-A.
PD 24-NOV-1998.
PF 06-JUN-1995; 471494.
PR 06-APR-1992; US-864475.
PR 05-APR-1991; US-681702.
PR 06-JUN-1995; US-471494.
PA (GENO) GEN HOSPITAL CORP.
PI Abou-Samra A, Juppner H, Kronenberg HM, Potts JT,
PI Schipani E, Segre GV;
DR WPI; 99-034124/03.
PT Antibody to parathyroid hormone receptor - for diagnostic or
PT therapeutic use
PS Claim 8; Column 19; 63pp; English.
CC This sequence is a fragment of a opossum parathyroid hormone (PTH)
CC receptor which is targeted by the antibody of the invention. The antibody
CC of the invention is immunoreactive with naturally occurring human, rat or
CC opossum PTH receptor. The antibody is useful for treating disorders
CC characterised by overstimulation of PTH receptors by their ligand and for
CC the diagnosis of PTH-related hypercalcaemia.
SQ Sequence 26 AA;

Query Match 64.0%; Score 48; DB 38; Length 26;
Best Local Similarity 57.1%; Pred. No. 2.94e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 11 cwdlssg 17
||:|:|
QY 120 CWNLVNG 126

Search completed: Thu Jul 8 18:24:29 1999
Job time : 18 secs.

W P R E F E R E N C E S

(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:23:36 1999; Maspar time 5.15 Seconds
69.958 Million cell updates/sec
Tabular output not generated.

Title: >US-09-041-236-2
Description: (120-128) from US09041236.pap (9 of 45)
Perfect Score: 75
Sequence: 1 CWNLVNGTV 9

Scoring table: PAM 150
Gap 11
Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.858; Variance 36.127; scale 0.660

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	58	77.3	291	2	DNA polymerase I homo	1.55e+00
2	57	76.0	80	1	hypothetical protein	2.35e+00
3	54	72.0	338	2	SAS2 protein - yeast	7.91e+00
4	54	72.0	708	2	killer toxin KHS prec	7.91e+00
5	53	70.7	134	2	cytochrome b5 homolog	1.18e+01
6	53	70.7	312	2	hypothetical protein	1.18e+01
7	53	70.7	330	2	endo-1,4-beta-xylanase	1.18e+01
8	53	70.7	451	2	probable D-alanine gl	1.18e+01
9	53	70.7	516	2	hypothetical protein	1.18e+01
10	52	69.3	323	2	cyclin homolog UME3 -	1.74e+01
11	52	69.3	453	2	hypothetical 49.4 kD	1.74e+01
12	52	69.3	491	2	hypothetical protein	1.74e+01
13	52	69.3	1150	2	myl protein - smut f	1.74e+01
14	51	68.0	356	1	gap junction protein	2.56e+01
15	51	68.0	443	2	X-Pro dipeptidase (EC	2.56e+01
16	51	68.0	564	2	Salp17R protein - vac	2.56e+01
17	51	68.0	564	2	A53R protein - vaccin	2.56e+01
18	51	68.0	581	2	probable aldehyde:fer	2.56e+01
19	50	66.7	378	2	quH protein - Emeric	3.75e+01
20	50	66.7	399	2	conserved hypotheticala	3.75e+01
21	50	66.7	412	2	probable intercellula	3.75e+01
22	50	66.7	575	2	acetylactate synthase	3.75e+01
23	50	66.7	659	2	hypothetical protein	3.75e+01

24	49	65.3	280	2	C70696	probable transport sy	5.46e+01
25	49	65.3	357	2	A49024	connexin40 - dog	5.46e+01
26	49	65.3	358	2	S23111	connexin 40 - mouse	5.46e+01
27	49	65.3	398	2	E64913	hypothetical protein	5.46e+01
28	49	65.3	477	2	JC2041	glucagon receptor pre	5.46e+01
29	49	65.3	485	2	JQ1957	glucagon receptor - r	5.46e+01
30	49	65.3	485	2	JC4363	glucagon receptor pre	5.46e+01
31	49	65.3	535	2	S48268	probable membrane pro	5.46e+01
32	49	65.3	550	2	A57519	parathyroid hormone r	5.46e+01
33	49	65.3	552	2	D64826	YbJD protein - Escher	5.46e+01
34	49	65.3	591	1	CB8Y2	L-lactate dehydrogena	5.46e+01
35	49	65.3	772	1	JQ2025	outer layer protein V	5.46e+01
36	49	65.3	772	1	JQ2024	outer layer protein V	5.46e+01
37	49	65.3	772	1	A44052	outer layer protein V	5.46e+01
38	49	65.3	1564	2	S55517	probable transport pr	5.46e+01
39	48	64.0	201	2	S37847	hypothetical protein	7.92e+01
40	48	64.0	471	2	S78597	argininosuccinate lya	7.92e+01
41	48	64.0	585	2	A39286	parathyroid hormone /	7.92e+01
42	48	64.0	591	2	I54195	parathyroid hormone/p	7.92e+01
43	48	64.0	591	2	S44203	parathyroid hormone-r	7.92e+01
44	48	64.0	593	2	A49191	parathyroid hormone/p	7.92e+01
45	48	64.0	896	2	B43817	transforming protein	7.92e+01

ALIGNMENTS

RESULT 1
ENTRY I64228 #type complete
TITLE DNA polymerase I homolog - Mycoplasma genitalium (SGC3)
ALTERNATE_NAMES hypothetical protein MG262; hypothetical protein MG468
ORGANISM #formal_name Mycoplasma genitalium
DATE 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
10-Oct-1997
ACCESSIONS I64228; G64251
REFERENCE A64200
#authors Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III, C.A.; Venter, J.C
#journal Science (1995) 270:397-403
#title The minimal gene complement of Mycoplasma genitalium.
#cross-references MUID:96026346
#accession I64228
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-291 ##label TIG1
##cross-references GB:U39705; GB:L43967; NID:g1045953; PID:g1045955; TIGR:MG262
#experimental_source strain G-37
#accession G64251
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-291 ##label TIG2
##cross-references GB:L43967; TIGR:MG468
#experimental_source strain G-37
GENETICS
#genetic_code SGC3
SUMMARY #length 291 #molecular-weight 33270 #checksum 3061
Query Match 77.3%; Score 58; DB 2; Length 291;
Best Local Similarity 55.6%; Pred. No. 1.55e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 45 CWNLIKAVY 53

Qy 120 CWNLVNGTV 128

RESULT 2

```

ENTRY          QOVZ10      #type complete
TITLE          hypothetical protein D-80 - vaccinia virus
ALTERNATE_NAMES D-ORF-E protein
ORGANISM       #formal_name vaccinia virus
DATE          04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change
              14-Nov-1997
ACCESSIONS    A03881; H42516
REFERENCE     A01146
#authors      Niles, E.G.; Condit, R.C.; Caro, P.; Davidson, K.; Matusick,
              L.; Seto, J.
#journal      Virology (1986) 153:96-112
#title        Nucleotide sequence and genetic map of the 16-kb vaccinia
              virus HindIII D fragment.
#cross-references MUID:86291159
#accession    A03881
#molecule_type DNA
#residues     1-80 #label N1L
#cross-references GB:M15058; NID:g335640; PID:g335649
#experimental_source strain WR
REFERENCE     A33172
#authors      Johnson, G.P.
#submission   submitted to GenBank, June 1990
#accession    H42516
#molecule_type DNA
#residues     1-80 #label JOH
#experimental_source strain Copenhagen
#length 80 #molecular-weight 9319 #checksum 756
SUMMARY

Query Match      76.0%; Score 57; DB 1; Length 80;
Best Local Similarity 66.7%; Pred. No. 2.35e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 39 CWKVVNGRV 47
   ||::||| |
Qy 120 CWNLVNGTV 128

RESULT 3
ENTRY  S48299      #type complete
TITLE  SAS2 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YK9553.03c; protein YMR126c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change
              12-Dec-1997
ACCESSIONS S48299; S53057
REFERENCE  S48299
#authors   Rivier, D.H.; Rine, J.; Errade, P.R.
#submission submitted to the EMBL Data Library, September 1994
#description SAS2, a gene involved in silencing HMR in yeast.
#accession  S48299
#molecule_type DNA
#residues  1-338 #label RIV
#cross-references EMBL:U14548; NID:g540231; PID:g540232
REFERENCE  S53055
#authors   Badcock, K.; Churcher, C.
#submission submitted to the EMBL Data Library, March 1995
#accession  S53057
#molecule_type DNA
#residues  1-338 #label BAD
#cross-references EMBL:248622; NID:g728663; PID:g728666; MIPS:YMR126c
GENETICS
#gene
#map_position 13R
SUMMARY #length 338 #molecular-weight 39206 #checksum 5167

Query Match      72.0%; Score 54; DB 2; Length 338;
Best Local Similarity 57.1%; Pred. No. 7.91e+00;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 262 CWHLIEG 268
   ||::||| |
Qy 120 CWNLVNG 126

```

```

RESULT 4
ENTRY  JQ1148      #type complete
TITLE  killer toxin KHS precursor - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES #formal_name Saccharomyces cerevisiae
ORGANISM 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
              14-Aug-1998
ACCESSIONS JQ1148
REFERENCE  JQ1148
#authors   Goto, K.; Fukuda, H.; Kichise, K.; Kitano, K.; Hara, S.
#journal   Agric. Biol. Chem. (1991) 55:1953-1958
#title     Cloning and nucleotide sequence of the KHS killer gene of
              Saccharomyces cerevisiae.
#cross-references MUID:92118315
#accession  JQ1148
#molecule_type DNA
#residues  1-708 #label GOT
#cross-references GB:S77712; NID:g243610; PID:g243611
#experimental_source strain No.115
GENETICS
#gene      KHS
#map_position 5
KEYWORDS   transmembrane protein
FEATURE    1-36
              37-708
SUMMARY    #domain signal sequence #status predicted #label SIG\
              #product KHS killer toxin #status predicted #label KHS
              #length 708 #molecular-weight 79534 #checksum 2859

Query Match      72.0%; Score 54; DB 2; Length 708;
Best Local Similarity 77.8%; Pred. No. 7.91e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 699 CWLVNGV 706
   ||::||| |
Qy 120 CWNLVNGTV 128

RESULT 5
ENTRY  T00796      #type complete
TITLE  cytochrome b5 homolog - Arabidopsis thaliana
ALTERNATE_NAMES F24L7.14 protein
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
DATE     12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
              05-Mar-1999
ACCESSIONS T00796
REFERENCE  Z14204
#authors   Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
              Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
              Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
              J.C.
#submission submitted to the EMBL Data Library, February 1998
#description Arabidopsis thaliana chromosome II BAC F24L7 genomic
              sequence.
#accession  T00796
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues  1-134 #label ROU
#cross-references EMBL:AC003974; NID:g2914688; PID:g2914701
#experimental_source cultivar Columbia
GENETICS
#map_position 2
#introns    29/3; 52/1
#note       F24L7.14
CLASSIFICATION #superfamily cytochrome b5; cytochrome b5 core homology
              #length 134 #molecular-weight 15016 #checksum 8729
SUMMARY

Query Match      70.7%; Score 53; DB 2; Length 134;
Best Local Similarity 55.6%; Pred. No. 1.18e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 22 CWIVNGKV 30
   ||::||| |

```

S26086

```
#authors      Hallick, R.B.
#submission   submitted to the EMBL Data Library, March 1992
#accession    S26088
#molecule_type DNA
#residues     1-516 ##label HAL
##cross-references EMBL:Z11874; NID:g14353; PID:e223926; PID:g1208513
GENETICS
#genome       chloroplast
#start_codon  TTG
#introns      17/2; 18/2; 222/3; 228/3
CLASSIFICATION
#superfamily Euglena gracilis chloroplast hypothetical
#protein 516
KEYWORDS      chloroplast
SUMMARY       #length 516 #molecular-weight 64333 #checksum 5374

Query Match      70.7%; Score 53; DB 2; Length 516;
Best Local Similarity 75.0%; Pred. No. 1.18e+01;
Matches          6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 15 WNLVDNTV 22
QY 121 WNLVNGTV 128
|||||
Query Match      70.7%; Score 52; DB 2; Length 323;
Best Local Similarity 50.0%; Pred. No. 1.74e+01;
Matches          4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 208 CWSLINDS 215
QY 120 CWNLVNGT 127
|||||

RESULT 10
ENTRY   S59373 #type complete
TITLE   cyclin homolog UME3 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YNL025c; RNA polymerase II holoenzyme
              cyclin-like chain; SSN8 protein
ORGANISM #formal_name Saccharomyces cerevisiae
DATE      30-Nov-1995 #sequence_revision 23-Feb-1996 #text_change
              17-Mar-1999
ACCESSIONS S59373; S59808; S62937; S62947; S60265
REFERENCE   S59373
#authors    Smith, J.B.; Mallory, M.J.; Strich, R.
#submission submitted to the EMBL Data Library, October 1994
#description UME3, a new C-type cyclin involved in the developmental
              switch between meiotic and mitotic cell division in S.
              cerevisiae.
#accession  S59373
#molecule_type DNA
#residues   1-323 ##label SMI
##cross-references EMBL:U10248; NID:g727250; PID:g727251
REFERENCE   S59808
#authors    Kuchin, S.; Yeghiayan, P.; Carlson, M.
#submission submitted to the EMBL Data Library, February 1995
#description Cyclin-dependent kinase and cyclin homologs, SSN3 and SSN8,
              contribute to transcriptional control in yeast.
#accession  S59808
#molecule_type DNA
#residues   1-323 ##label KUC
##cross-references EMBL:U20635; NID:g676866; PID:g676867
REFERENCE   S62920
#authors    Andre, B.; Iraqui Houssaini, I.; Urrestarazu, L.A.; Vissers,
              S.
#submission submitted to the Protein Sequence Database, April 1996
#accession  S62937
#molecule_type DNA
#residues   1-323 ##label AND
##cross-references EMBL:Z71301; NID:g1301858; PID:e239871; PID:g1301859;
              MIPS:YNL025C
#experimental_source strain S288C
REFERENCE   S62944
#authors    Duesterhoeft, A.; Floeth, M.; Fritz, C.; Heuss-Neitzel, D.;
              Hilbert, H.; Moestl, D.
#submission submitted to the Protein Sequence Database, April 1996
#accession  S62947
#molecule_type DNA
#residues   1-323 ##label DUE
##cross-references EMBL:Z71301; NID:g1301858; PID:e239871; PID:g1301859;
              MIPS:YNL025C
#experimental_source strain S288C
REFERENCE   S60264
#authors    Smith, J.B.; Mallory, M.J.; Strich, R.
#submission submitted to the EMBL Data Library, October 1994
#description UME3, a new C-type cyclin involved in the developmental
              switch between meiotic and mitotic cell division in S.
              cerevisiae.
#accession  S59373
#molecule_type DNA
#residues   1-323 ##label SMI
##cross-references EMBL:U10248; NID:g727250; PID:g727251
REFERENCE   S59808
#authors    Kuchin, S.; Yeghiayan, P.; Carlson, M.
#submission submitted to the EMBL Data Library, February 1995
#description Cyclin-dependent kinase and cyclin homologs, SSN3 and SSN8,
              contribute to transcriptional control in yeast.
#accession  S59808
#molecule_type DNA
#residues   1-323 ##label KUC
##cross-references EMBL:U20635; NID:g676866; PID:g676867
REFERENCE   S62920
#authors    Andre, B.; Iraqui Houssaini, I.; Urrestarazu, L.A.; Vissers,
              S.
#submission submitted to the Protein Sequence Database, April 1996
#accession  S62937
#molecule_type DNA
#residues   1-323 ##label AND
##cross-references EMBL:Z71301; NID:g1301858; PID:e239871; PID:g1301859;
              MIPS:YNL025C
#experimental_source strain S288C
REFERENCE   S62944
#authors    Duesterhoeft, A.; Floeth, M.; Fritz, C.; Heuss-Neitzel, D.;
              Hilbert, H.; Moestl, D.
#submission submitted to the Protein Sequence Database, April 1996
#accession  S62947
#molecule_type DNA
#residues   1-323 ##label DUE
##cross-references EMBL:Z71301; NID:g1301858; PID:e239871; PID:g1301859;
              MIPS:YNL025C
#experimental_source strain S288C
```

```
#authors      Liao, S.M.; Zhang, J.; Jeffery, D.A.; Koleske, A.J.;
              Thompson, C.M.; Chao, D.M.; Viljoen, M.; van Vuuren,
              H.J.J.; Young, R.A.
#journal      Nature (1995) 374:193-196
#title        A kinase-cyclin pair in the RNA polymerase II holoenzyme.
#cross-references MUID:95183147
#accession    S60265
#status       nucleic acid sequence not shown
#molecule_type DNA
#residues     1-323 ##label LIA
##cross-references EMBL:U20221; NID:g699505; PID:g699506
GENETICS
#gene         SGD:SSN8; UME3; SRB11
##cross-references SGD:S0004970; MIPS:YNL025c
#map_position 14L
KEYWORDS      transmembrane protein
FEATURE       222-238
SUMMARY       #length 323 #molecular-weight 37790 #checksum 9639

Query Match      69.3%; Score 52; DB 2; Length 323;
Best Local Similarity 50.0%; Pred. No. 1.74e+01;
Matches          4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 208 CWSLINDS 215
QY 120 CWNLVNGT 127
|||||

RESULT 11
ENTRY   F65250 #type complete
TITLE   hypothetical 49.4 kD protein in tsr-mdb intergenic region -
              Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE      12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
              20-Mar-1998
ACCESSIONS F65250; S56582
REFERENCE   A64720
#authors    Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
              Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
              Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
              Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
              Y.
#journal     Science (1997) 277:1453-1462
#title       The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession   F65250
#status       nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues     1-453 ##label BLAT
##cross-references GB:AE000506; GB:U00096; NID:g2367377; PID:g2367379;
              UWGP:b4356
#experimental_source strain K-12, substrain MG1655
REFERENCE   S56314
#authors    Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.;
              Blattner, F.R.
#journal     Nucleic Acids Res. (1995) 23:2105-2119
#title       Analysis of the Escherichia coli genome VI: DNA sequence of
              the region from 92.8 through 100 minutes.
#cross-references MUID:95334362
#accession   S56582
#status       nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues     262-453 ##label BUR
##cross-references EMBL:U14003; NID:g1263172; PID:g537198
#note        the nucleotide sequence was submitted to the EMBL Data
              Library, August 1994
CLASSIFICATION #superfamily hexuronate transporter
SUMMARY        #length 453 #molecular-weight 49440 #checksum 2337

Query Match      69.3%; Score 52; DB 2; Length 453;
Best Local Similarity 44.4%; Pred. No. 1.74e+01;
Matches          4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

```
Db 376 CWGLIHVAV 384
|||:::|
QY 120 CWNLVNGTV 128

RESULT 12
ENTRY S73110 #type complete
TITLE hypothetical protein 491 - red alga (Porphyra purpurea)
ALTERNATE_NAMES chloroplast
ORGANISM #formal_name chloroplast Porphyra purpurea
DATE 19-Mar-1997 #sequence_revision 09-May-1997 #text_change
13-Sep-1998
ACCESSIONS S73110
REFERENCE S73108
#authors Reith, M.; Munholland, J.
#journal Plant Mol. Biol. Rep. (1995) 13:333-335
#title Complete nucleotide sequence of the Porphyra purpurea
chloroplast genome.
#accession S73110
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-491 ##label REI
##cross-references EMBL:U38804; NID:gl276652; PID:gl276655
##note the nucleotide sequence was submitted to the EMBL Data
Library, October 1995
GENETICS
#genome chloroplast
#superfamily hypothetical protein slr0480;
CLASSIFICATION FtsH/SEC18/CDC48-type ATP-binding domain homology
KEYWORDS chloroplast; P-loop
FEATURE
242-447 #domain FtsH/SEC18/CDC48-type ATP-binding domain
homology #label VAMP\
266-273 #region nucleotide-binding motif A (P-loop)
#length 491 #molecular-weight 56030 #checksum 3792
SUMMARY
Query Match 69.3%; Score 52; DB 2; Length 491;
Best Local Similarity 57.1%; Pred. No. 1.74e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 46 CWDFVDG 52
|||:::|
QY 120 CWNLVNGTV 126

RESULT 13
ENTRY S58775 #type complete
TITLE mypl protein - smut fungus (Ustilago maydis)
ALTERNATE_NAMES #formal_name Ustilago maydis #common_name corn smut
ORGANISM 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change
05-Jun-1998
ACCESSIONS S58775
REFERENCE S58775
#authors Giasson, L.; Kronstad, J.W.
#journal Genetics (1995) 141:491-501
#title Mutations in the mypl gene of Ustilago maydis attenuate
mycelial growth and virulence.
#accession S58775
#status preliminary
#molecule_type DNA
##residues 1-1150 ##label GIA
##cross-references EMBL:L33919; NID:g886415; PID:g886416
GENETICS
#genome mypl
SUMMARY
#length 1150 #molecular-weight 123387 #checksum 61
Query Match 69.3%; Score 52; DB 2; Length 1150;
Best Local Similarity 66.7%; Pred. No. 1.74e+01;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 822 CWPLRVTV 830
```

```
QY 120 CWNLVNGTV 128
|||:::|

RESULT 14
ENTRY A42053 #type complete
TITLE gap junction protein Cx40 - rat
ALTERNATE_NAMES connexin 40
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 04-Mar-1993 #sequence_revision 02-Jun-1994 #text_change
05-Sep-1997
ACCESSIONS A42053
REFERENCE A42053
#authors Haefliger, J.A.; Bruzzone, R.; Jenkins, N.A.; Gilbert, D.J.;
Copeland, N.G.; Paul, D.L.
#journal J. Biol. Chem. (1992) 267:2057-2064
#title Four novel members of the connexin family of gap junction
proteins. Molecular cloning, expression, and chromosome
mapping.
#cross-references MUID:92112940
#accession A42053
##molecule_type DNA
##residues 1-356 ##label HAE
##cross-references GB:M76535; NID:g203667; PID:g203668
##note sequence extracted from NCBI backbone (NCBIP:76097)
CLASSIFICATION #superfamily gap junction protein
KEYWORDS gap junction; phosphoprotein; transmembrane protein
FEATURE
1-23 #domain intracellular #status predicted #label INT1\
24-41 #domain transmembrane #status predicted #label TM1\
42-77 #domain extracellular #status predicted #label EE1\
78-97 #domain transmembrane #status predicted #label TM2\
98-148 #domain intracellular #status predicted #label INT2\
149-185 #domain transmembrane #status predicted #label EE2\
186-205 #domain extracellular #status predicted #label TM3\
206-232 #domain transmembrane #status predicted #label TM4\
233-356 #domain intracellular #status predicted #label INT3
SUMMARY
#length 356 #molecular-weight 40233 #checksum 3566
Query Match 68.0%; Score 51; DB 1; Length 356;
Best Local Similarity 55.6%; Pred. No. 2.56e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 134 CWKEVNGKI 142
|||:::|
QY 120 CWNLVNGTV 128

RESULT 15
ENTRY H65189 #type complete
TITLE X-Pro dipeptidase (EC 3.4.13.9) - Escherichia coli (strain
K-12)
ALTERNATE_NAMES ininodipeptidase; prolidase; proline dipeptidase
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
05-Dec-1997
ACCESSIONS H65189; S30738; JQ0753
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession H65189
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-443 ##label BLAT
##cross-references GB:AE000460; GB:U000096; NID:g2367315; PID:gl790282;
UWGP:b3847
```

```
#experimental_source strain K-12, Substrain MG1655
REFERENCE S30660
#authors Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
#journal Science (1992) 257:771-778
#title Analysis of the Escherichia coli genome: DNA sequence of the
#cross-references MUID:92358234 region from 84.5 to 86.5 minutes.
#accession S30738
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-22,'X','24-120,'X',122-443 #label DAN
#cross-references EMBL:M87049; NID:g836656; PID:g148247
#note the nucleotide sequence was submitted to the EMBL Data
Library, November 1992

REFERENCE JQ0753
#authors Nakahigashi, K.; Inokuchi, H.
#journal Nucleic Acids Res. (1990) 18:6439
#title Nucleotide sequence between the fadB gene and the rrnA operon
from Escherichia coli.
#cross-references MUID:91057145
#accession JQ0753
#status significant sequence differences
#molecule_type DNA
#cross-references EMBL:X54687

GENETICS
#gene pepQ
#map_position 86 min
KEYWORDS dipeptide hydrolase; hydrolase
SUMMARY #length 443 #molecular-weight 50176 #checksum 2467

Query Match 68.0%; Score 51; DB 2; Length 443;
Best Local Similarity 71.4%; Pred. NO. 2.56e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 68 CWLLVDG 74
QY 120 CWNLVNG 126

Search completed: Thu Jul 8 18:23:50 1999
Job time : 14 secs.
```

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:22:32 1999; Maspar time 3.26 Seconds
Tabular output not generated. 78.057 Million cell updates/sec

Title: >US-09-041-236-2
Description: (120-128) from US09041236.pap (9 of 45)
Perfect Score: 75
Sequence: 1 CWNLVNGTV 9

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 24.590; Variance 33.675; scale 0.730

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description	Pred. No.
1	58	77.3	1	EX53_MYCGE	6.30e+01
2	57	76.0	1	YVDE_VACCV	9.84e+01
3	56	74.7	600	DHSA_PARDE	1.53e+00
4	55	73.3	288	MYO2_HUMAN	2.37e+00
5	54	72.0	338	SAS2_PROTEIN	3.64e+00
6	54	72.0	708	KILLER_TOXIN_KHS_PRCU	3.64e+00
7	53	70.7	312	XYN4_CALSA	5.57e+00
8	53	70.7	312	OLF2_CHICK	5.57e+00
9	53	70.7	312	OLF4_CHICK	5.57e+00
10	53	70.7	312	OLF6_CHICK	5.57e+00
11	53	70.7	312	OLF5_CHICK	5.57e+00
12	53	70.7	330	XYN2_BACST	5.57e+00
13	53	70.7	516	YCX4_EUGGR	5.57e+00
14	52	69.3	323	UME3_YEAST	8.49e+00
15	52	69.3	453	YJUL_ECOLI	8.49e+00
16	52	69.3	491	YC46_PORPU	8.49e+00
17	51	68.0	355	1	1.88e+01
18	51	68.0	443	PEPO_ECOLI	1.28e+01
19	51	68.0	564	1	1.28e+01
20	51	68.0	564	1	1.28e+01
21	50	66.7	399	1	1.93e+01
22	50	66.7	559	1	1.93e+01
23	49	65.3	356	1	2.89e+01

24	49	65.3	357	1	CXA5_MOUSE	GAP JUNCTION ALPHA-5 P	2.89e+01
25	49	65.3	373	1	NOLL_RHILO	MODULATION PROTEIN NOL	2.89e+01
26	49	65.3	477	1	GLR_HUMAN	GLUCAGON RECEPTOR PREC	2.89e+01
27	49	65.3	485	1	GLR_RAT	GLUCAGON RECEPTOR PREC	2.89e+01
28	49	65.3	485	1	GLR_MOUSE	GLUCAGON RECEPTOR PREC	2.89e+01
29	49	65.3	535	1	YBV3_YEAST	HYPOTHETICAL 59.1 KD P	2.89e+01
30	49	65.3	550	1	PTR2_HUMAN	PARATHYROID HORMONE RE	2.89e+01
31	49	65.3	552	1	YBJD_ECOLI	HYPOTHETICAL 63.6 KD P	2.89e+01
32	49	65.3	591	1	CYB2_YEAST	CYTOCHROME B2 PRECURSO	2.89e+01
33	49	65.3	772	1	VP4_ROTBA	OUTER CAPSID PROTEIN V	2.89e+01
34	49	65.3	772	1	VP4_ROTBB	OUTER CAPSID PROTEIN V	2.89e+01
35	49	65.3	1564	1	PDBA_YEAST	ATP-DEPENDENT PERMEASE	2.89e+01
36	48	64.0	134	1	CYB5_BRAOL	CYTOCHROME B5	4.30e+01
37	48	64.0	201	1	YKD0_YEAST	HYPOTHETICAL 23.0 KD P	4.30e+01
38	48	64.0	261	1	PCNA_RAT	PROLIFERATING CELL NUC	4.30e+01
39	48	64.0	490	1	AMPI_STRLI	XAA-PRO AMINOPEPTIDASE	4.30e+01
40	48	64.0	585	1	PTRR_DIDMA	PARATHYROID HORMONE/PA	4.30e+01
41	48	64.0	591	1	PTRR_RAT	PARATHYROID HORMONE/PA	4.30e+01
42	48	64.0	619	1	NTDOL_RAT	SODIUM-DEPENDENT DOPAM	4.30e+01
43	48	64.0	776	1	CHSL_CANAL	CHITIN SYNTHASE I (EC	4.30e+01
44	48	64.0	896	1	CBL_MOUSE	PROTO-ONCOGENE C-CBL	4.30e+01
45	48	64.0	2110	1	MCAS_MYCBO	MYCOCEROSIC ACID SYNTH	4.30e+01

ALIGNMENTS

RESULT	ID	EX53_MYCGE	STANDARD	PRT	291 AA
AC	Q49406				
DT	01-NOV-1997	(REL. 35, CREATED)			
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)			
DE	POTENTIAL 5'-3' EXONUCLEASE (EC 3.1.11.-).				
GN	POLA OR MG262				
OS	MYCOPLASMA GENITALIUM				
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;				
OC	MYCOPLASMATACEAE; MYCOPLASMA				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 33530 / G-37;				
RX	MEDLINE; 96026346				
RA	FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,				
RA	FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,				
RA	FRITCHMAN J.D., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.D.,				
RA	NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,				
RA	TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,				
RA	PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;				
RT	"The minimal gene complement of Mycoplasma genitalium.";				
RL	SCIENCE 270:397-403(1995)				
CC	-1- SIMILARITY: HIGH, TO THE EXONUCLEASE DOMAIN OF SOME BACTERIAL				
CC	DNA POLYMERASES OF FAMILY A.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U39705; G1045955; -				
DR	TIGR; MG262; -				
KW	HYDROLASE; NUCLEASE; EXONUCLEASE; DNA-BINDING.				
SK	SEQUENCE 291 AA; 33270 MW; F94D3673 CRC32;				

Query Match 77.3%; Score 58; DB 1; Length 291;
Best Local Similarity 55.6%; Pred. No. 6.30e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 45 CWNLIKAV 53
I I I I I I I I I I
Qy 120 CWNLVNGTV 128


```
CC EMBL: AF014398; G2406666; -
DR PROSITE: PS00629; IMP_1; 1.
DR PROSITE: PS00630; IMP_2; 1.
DR PIR: PF00459; Inositol_P; 1.
KW HYDROLASE.
SQ SEQUENCE 288 AA; 31321 MW; 8EFC7AA1 CRC32;

Query Match 73.3%; Score 55; DB 1; Length 288;
Best Local Similarity 55.6%; Pred. No. 2.37e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 229 CWDLAATV 237
||:|:|
QY 120 CWNLVNGIV 128

RESULT 5
ID SAS2_YEAST STANDARD; PRT; 338 AA.
AC P40963;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SAS2 PROTEIN.
GN SAS2 OR ESO1 OR YMR127C OR YM9553.03C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETEALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA RIVIER D.H., RINE J., ERRADE P.R.;
RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA BADCOCK K., CHURCHER C., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE; 96376969.
RA REIFSNYDER C., LOWELL J., CLARKE A., PILLUS L.;
RL NAT. GENET. 16:109-109(1997).
CC -1- FUNCTION: PLAYS A ROLE IN TELOMERIC SILENCING AND G0 CONTROL.
CC -1- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U14548; G540232; -
DR EMBL: 248622; G728666; -
DR PIR: S48299; S48299.
DR SGD: L0001802; SAS2.
KW ZINC-FINGER.
FT ZN-FING 101 123 C2HC-TYPE.
FT SIMILAR 206 229 TO SOME ACETYLTRANSFERASES.
SQ SEQUENCE 338 AA; 39206 MW; CSCB1BD8 CRC32;

Query Match 72.0%; Score 54; DB 1; Length 338;
Best Local Similarity 57.1%; Pred. No. 3.64e+00;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 262 CWHLIEG 268
||:|:|
QY 120 CWNLVNG 126

RESULT 6
ID KHS1_YEAST STANDARD; PRT; 708 AA.
AC P39690;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE KILLER TOXIN KHS PRECURSOR (KILLER OF HEAT SENSITIVE).
GN KHS1 OR KHS.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETEALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 37-42.
RX STRAIN=115;
RX MEDLINE; 92118315.
RA GOTO K., FUKUDA H., KICHISE K., KITANO K., HARA S.;
RL "Cloning and nucleotide sequence of the KHS killer gene of
RL Saccharomyces cerevisiae."
CC -1- FUNCTION: KILL SENSITIVE STRAINS OF YEAST. OPTIMAL PH IS 4-4.5
CC AND IT IS UNSTABLE ABOVE 30 DEGREES CELSIUS.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: TO YEAST YER187W.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: S77712; G243611; -
DR PIR: J01148; J01148.
DR SGD: L0000900; KHS1.
KW TOXIN; SIGNAL; TRANSMEMBRANE.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 36 KILLER TOXIN KHS.
FT CHAIN 37 708 POTENTIAL.
FT TRANSMEM 77 97 POTENTIAL.
FT TRANSMEM 380 400 POTENTIAL.
FT TRANSMEM 466 486 POTENTIAL.
SQ SEQUENCE 708 AA; 79535 MW; CE8A97FB CRC32;

Query Match 72.0%; Score 54; DB 1; Length 708;
Best Local Similarity 77.8%; Pred. No. 3.64e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 699 CW-LVNGV 706
||:|:|:|
QY 120 CWNLVNGIV 128

RESULT 7
ID XYN4_CALSA STANDARD; PRT; 312 AA.
AC P23557;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PUTATIVE ENDO-1,4-BETA-XYLANASE (EC 3.2.1.8) (XYLANASE) (1,4-BETA-D-
DE XYLAN XYLANOXYDOLASE) (ORF4).
OS CALDOCELLUM SACCHAROLYTICUM (CALDICELLULOSITRUPTOR SACCHAROLYTICUS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP;
OC THERMOANAEROBACTER GROUP; CALDICELLULOSITRUPTOR.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90253140.
RA LUETHI E., LOVE D.R., MCANULTY J., WALLACE C., CAUGHEY P.A.,
```

RA SAUL D., BERQUIST P.L.;
 RT "Cloning, sequence analysis, and expression of genes encoding xylan-
 RT degrading enzymes from the thermophile 'Caldocellum
 RL saccharolyticum'.
 RL APPL. ENVIRON. MICROBIOL. 56:1017-1024(1990).
 CC -!- FUNCTION: COULD BE A XYLANASE (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
 CC LINKAGES IN XYLANS.
 CC -!- PATHWAY: XYLAN DEGRADATION.
 CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M34459; G144299; -
 DR PIR: D37202; D37202.
 DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
 DR PFAM; PF00331; glycosyl_hydro3; 1.
 DR HSP; P10478; IXYZ.
 DR XLAN DEGRADATION; HYDROLASE; GLYCOSIDASE.
 FT DOMAIN 23 274 CATALYTIC (POTENTIAL).
 FT ACT_SITE 104 104 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 216 216 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 312 AA; 36494 MW; F9C3F64F CRC32;

Query Match 70.7%; Score 53; DB 1; Length 312;
 Best Local Similarity 44.4%; Pred. No. 5.57e+00;
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 98 CWDVNEAI 106
 QY 120 CWNLVNGTV 128
 ||:|:|
 ||:|:|

RESULT 8
 ID OLUF4_CHICK STANDARD; PRT; 312 AA.
 AC P37068;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE OLFACTORY RECEPTOR-LIKE PROTEIN COR2.
 GN COR2.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
 [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=OLFACTORY EPITHELIUM;
 RX MEDLINE; 96317247.
 RA NEF S., ALLAMAN I., FIUMELLI H., DE CASTRO E., NEF P.;
 RT "Olfaction in birds: differential embryonic expression of nine
 RT putative odorant receptor genes in the avian olfactory system.";
 RL MECH. DEV. 55:65-77(1996).
 CC -!- FUNCTION: PUTATIVE ODORANT RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z79588; E262002; -
 DR GCRDB; GCR_1099; -

DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
 DR PFAM; PF00001; 7tm_1; 1.
 KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
 KW OLFACTION; MULTIGENE FAMILY.
 FT DOMAIN 1 26 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 27 49 1 (POTENTIAL).
 FT DOMAIN 50 57 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 58 79 2 (POTENTIAL).
 FT DOMAIN 80 100 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 101 120 3 (POTENTIAL).
 FT DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 140 184 4 (POTENTIAL).
 FT DOMAIN 165 205 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 206 226 5 (POTENTIAL).
 FT DOMAIN 227 239 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 240 260 6 (POTENTIAL).
 FT DOMAIN 261 271 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 272 292 7 (POTENTIAL).
 FT DOMAIN 293 312 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 5 5 POTENTIAL.
 FT DISULFID 97 179 BY SIMILARITY.
 SQ SEQUENCE 312 AA; 35255 MW; 58164319 CRC32;

Query Match 70.7%; Score 53; DB 1; Length 312;
 Best Local Similarity 71.4%; Pred. No. 5.57e+00;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 141 CWELVKS 147
 QY 120 CWNLVNG 126
 ||:|:|
 ||:|:|

RESULT 9
 ID OLUF4_CHICK STANDARD; PRT; 312 AA.
 AC P37070;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE OLFACTORY RECEPTOR-LIKE PROTEIN COR4.
 GN COR4.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
 [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=OLFACTORY EPITHELIUM;
 RX MEDLINE; 96317247.
 RA NEF S., ALLAMAN I., FIUMELLI H., DE CASTRO E., NEF P.;
 RT "Olfaction in birds: differential embryonic expression of nine
 RT putative odorant receptor genes in the avian olfactory system.";
 RL MECH. DEV. 55:65-77(1996).
 CC -!- FUNCTION: PUTATIVE ODORANT RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z79593; E262006; -
 DR GCRDB; GCR_1095; -
 DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
 DR PFAM; PF00001; 7tm_1; 1.
 KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
 KW OLFACTION; MULTIGENE FAMILY.
 FT DOMAIN 1 26 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 27 49 1 (POTENTIAL).
 FT DOMAIN 50 57 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 58 79 2 (POTENTIAL).

FT DOMAIN 80 100 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 101 120 3 (POTENTIAL).
FT DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 164 4 (POTENTIAL).
FT DOMAIN 165 205 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 206 226 5 (POTENTIAL).
FT DOMAIN 227 239 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 240 260 6 (POTENTIAL).
FT DOMAIN 261 271 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 272 292 7 (POTENTIAL).
FT DOMAIN 293 312 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 5 5 POTENTIAL.
FT DISULFID 97 179 BY SIMILARITY.
SQ SEQUENCE 312 AA; 35093 MW; 252C5E40 CRC32;

Query Match 70.7%; Score 53; DB 1; Length 312;
Best Local Similarity 71.4%; Pred. No. 5.57e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 141 CWRLVK 147
QY 120 CWNLVNG 126

RESULT 10
ID OLF6-CHICK STANDARD; PRT; 312 AA.
AC P37072; Q98912;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE OLFATORY RECEPTOR-LIKE PROTEIN COR6.
GN COR6.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=OLFATORY EPITHELIUM;
RX MEDLINE; 96317247.
RA NEF S., ALLAMAN I., FIUMELLI H., DE CASTRO E., NEF P.;
RT "Olfaction in birds: differential embryonic expression of nine
putative odorant receptor genes in the avian olfactory system.";
RL MECH. DEV. 55:65-77(1996).
CC -!- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----

CC EMBL; Z79590; E262053; -
CC GCRDB; GCR_1092; -
CC DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; FALSE_NEG.
CC DR PFAM; PF00001; 7tm1; 1.
CC G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
CC OLFACATION; MULTIGENE FAMILY.
FT DOMAIN 1 26 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 27 49 1 (POTENTIAL).
FT DOMAIN 50 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 79 2 (POTENTIAL).
FT DOMAIN 80 100 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 101 120 3 (POTENTIAL).
FT DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 164 4 (POTENTIAL).
FT DOMAIN 165 205 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 206 226 5 (POTENTIAL).
FT DOMAIN 227 239 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 240 260 6 (POTENTIAL).

FT DOMAIN 261 271 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 272 292 7 (POTENTIAL).
FT DOMAIN 293 312 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 5 5 POTENTIAL.
FT DISULFID 97 179 BY SIMILARITY.
SQ SEQUENCE 312 AA; 35179 MW; A7FD57D CRC32;

Query Match 70.7%; Score 53; DB 1; Length 312;
Best Local Similarity 71.4%; Pred. No. 5.57e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 141 CWRLVK 147
QY 120 CWNLVNG 126

RESULT 11
ID OLF5-CHICK STANDARD; PRT; 312 AA.
AC P37071;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE OLFATORY RECEPTOR-LIKE PROTEIN COR5.
GN COR5.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OLFATORY EPITHELIUM;
RX MEDLINE; 96317247.
RA NEF S., ALLAMAN I., FIUMELLI H., DE CASTRO E., NEF P.;
RT "Olfaction in birds: differential embryonic expression of nine
putative odorant receptor genes in the avian olfactory system.";
RL MECH. DEV. 55:65-77(1996).
CC -!- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----

CC EMBL; Z79589; E262003; -
CC GCRDB; GCR_1093; -
CC DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
CC DR PFAM; PF00001; 7tm1; 1.
CC G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
CC OLFACATION; MULTIGENE FAMILY.
FT DOMAIN 1 26 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 27 49 1 (POTENTIAL).
FT DOMAIN 50 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 79 2 (POTENTIAL).
FT DOMAIN 80 100 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 101 120 3 (POTENTIAL).
FT DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 164 4 (POTENTIAL).
FT DOMAIN 165 205 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 206 226 5 (POTENTIAL).
FT DOMAIN 227 239 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 240 260 6 (POTENTIAL).
FT DOMAIN 261 271 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 272 292 7 (POTENTIAL).
FT DOMAIN 293 312 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 5 5 POTENTIAL.
FT DISULFID 97 179 BY SIMILARITY.
SQ SEQUENCE 312 AA; 35256 MW; 523CB56C CRC32;

Query Match 70.7%; Score 53; DB 1; Length 312;
Best Local Similarity 71.4%; Pred. No. 5.57e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 141 CWRLVK 147
QY 120 CWNLVNG 126

RESULT 11
ID OLF5-CHICK STANDARD; PRT; 312 AA.
AC P37071;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE OLFATORY RECEPTOR-LIKE PROTEIN COR5.
GN COR5.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OLFATORY EPITHELIUM;
RX MEDLINE; 96317247.
RA NEF S., ALLAMAN I., FIUMELLI H., DE CASTRO E., NEF P.;
RT "Olfaction in birds: differential embryonic expression of nine
putative odorant receptor genes in the avian olfactory system.";
RL MECH. DEV. 55:65-77(1996).
CC -!- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----

Query Match 70.7%; Score 53; DB 1; Length 312;

Best Local Similarity 71.4%; Pred. No. 5.57e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 141 CWRLVEG 147
QY 120 CWNLYNG 126

RESULT 12
ID XYN2 BACST STANDARD; PRT; 330 AA.
AC P45703;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE ENDO-1.4-BETA-XYLANASE PRECURSOR (EC 3.2.1.8) (XYLANASE)
DE (1.4-BETA-D-XLAN XYLANOXYDOLASE).
GN XYN.
OS BACILLUS THERMOPHILUS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 3-7.
RC STRAIN-NO.21;
RX MEDLINE; 94354640.
RA BABA T., SHINKE R., NANNORI T.;
RT "Identification and characterization of clustered genes for
thermostable xylan-degrading enzymes, beta-xylosidase and xylanase,
of Bacillus stearothermophilus 21";
RT APPL. ENVIRON. MICROBIOL. 60:2252-2258(1994).
RL -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
LINKAGES IN XYLANS.
CC -!- PATHWAY: XLAN DEGRADATION.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- INDUCTION: BY XLOSE AND XLAN.
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: D28121; G499714; -;
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PFAM; PF00331; glycosyl_hydro3; 1.
DR HSP; P10478; 1XYZ.
KW XLAN DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
FT SIGNAL 1 ?
FT CHAIN ?
FT ACT_SITE 133 330 XYLANASE.
FT ACT_SITE 133 133 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 240 240 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 330 AA; 38473 MW; 3FAE0336 CRC32;
Query Match 70.7%; Score 53; DB 1; Length 330;
Best Local Similarity 44.4%; Pred. No. 5.57e+00;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 127 CWDVINEAV 135
QY 120 CWNLYNGTV 128

RESULT 13
ID YCX4_EUGGR STANDARD; PRT; 516 AA.
AC P30397;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 64.3 KD PROTEIN IN RPS 3' REGION (ORF516).
OS EUGLA GRACILIS.

CHLOROPLAST.
EUKARYOTA; EUGLENOZOEA; EUGLENIDA; EUGLENALES; EUGLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-2;
RX MEDLINE; 93347989.
RA HALLICK R.B., HONG L., DRAGER R.G., FAVREAU M.R., MONFORT A.,
RA ORSAT B., SPIELMANN A., STUTZ E.;
RT "Complete sequence of Euglena gracilis chloroplast DNA";
RL NUCLEIC ACIDS RES. 21:3537-3544(1993).
CC -!- SIMILARITY: TO GROUP II INTRON MATURASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: X70810; E1204055; -;
DR EMBL; Z11874; E1204062; -;
DR PIR; S26088; S26088.
DR PIR; S34525; S34525.
KW CHLOROPLAST; HYPOTHETICAL PROTEIN.
SQ SEQUENCE 516 AA; 64333 MW; CE19DE9F CRC32;
Query Match 70.7%; Score 53; DB 1; Length 516;
Best Local Similarity 75.0%; Pred. No. 5.57e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 15 WNLVDNTV 22
QY 121 WNLVNGTV 128

RESULT 14
ID UME3_YEAST STANDARD; PRT; 323 AA.
AC P47821;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE II HOLOENZYME CYCLIN-LIKE SUBUNIT.
GN UME3 OR SSN8 OR SRB1 OR YNL025C OR N2805.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE; 95183147.
RA LIAO S.-M., ZHANG J., JEFFREY D.A., KOLESKA A.J., THOMPSON C.M.,
RA CHAO D.M., VILJOEN M., VAN VUUREN H.J.J., YOUNG R.A.;
RT "A kinase-cyclin pair in the RNA polymerase II holoenzyme";
RL NATURE 374:193-196(1995).
CC [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RA SMITH J.B., MALLORY M.J., STRICH R.;
RL SUBMITTED (XXX-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE; 95249601.
RA KUCHIN S., YEGHIAYAN P., CARLSON M.;
RT "Cyclin-dependent protein kinase and cyclin homologs SSN3 and SSN8
contribute to transcriptional control in yeast";
RL PROC. NATL. ACAD. SCI. U.S.A. 92:4006-4010(1995).
CC [4]
RP SEQUENCE FROM N.A.
RA DUESTERHOEFFT A., FLOETH M., FRITZ C., HEUSS-NEITZEL D., HILBERT H.,
RA MOESTL D.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

THIS PAGE BLANK (USPTO)



Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:23:01 1999; MasPar time 8.43 Seconds
Tabular output not generated.
58.242 Million cell updates/sec

Title: >US-09-041-236-2
Description: (120-128) from US09041236.pep (9 of 45)
Perfect Score: 75
Sequence: 1 CWNLVNGTV 9

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrmb19
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 24.124; Variance 32.855; scale 0.734

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	75	100.0	075326	SEMAPHORIN L.	2.45e-04
2	65	86.7	088371	SEMAPHORIN L. (FRAGMENT)	3.70e-02
3	58	77.3	152	NTP PYROPHOSPHODIOLASE	9.99e-01
4	54	73.3	367	ENDO-1,4-BETA-XYLANASE	3.84e+00
5	54	72.0	308	F49A5.4 PROTEIN.	5.96e+00
6	54	72.0	539	T05H4.9 PROTEIN.	5.96e+00
7	53	70.7	134	PUTATIVE CYTOCHROME B5	9.20e+00
8	53	70.7	312	OLFACTORY RECEPTOR 3 (9.20e+00
9	53	70.7	451	D-ALANINE GLYCINE PERM	9.20e+00
10	53	69.3	132	CYTOTOXIC B5	1.41e+01
11	52	69.3	132	CYTOTOXIC B5	1.41e+01
12	52	69.3	317	T25E12.9 PROTEIN.	1.41e+01
13	52	69.3	134	HYPOTHETICAL 58.7 KD P	1.41e+01
14	52	69.3	536	BETA-XYLAN ENDOHYDROLA	1.41e+01
15	52	69.3	1150	SMUT FUNGUS (CLONE PM1	1.41e+01
16	51	68.0	134	CYTOTOXIC B5 (FRAGMENT	2.16e+01
17	51	68.0	212	GP120, V1 TO V5 REGION	2.16e+01
18	51	68.0	212	GP120, V1 TO V5 REGION	2.16e+01
19	51	68.0	212	GP120, V1 TO V5 REGION	2.16e+01
20	51	68.0	212	GP120, V1 TO V5 REGION	2.16e+01

21	51	68.0	212	14	075317	GP120, V1 TO V5 REGION	2.16e+01
22	51	68.0	212	14	075316	GP120, V1 TO V5 REGION	2.16e+01
23	51	68.0	322	3	074237	XYLOSE REDUCTASE (EC 1	2.16e+01
24	51	68.0	448	5	017610	C28D4.2.	2.16e+01
25	51	68.0	556	8	095749	RIBOSOMAL PROTEIN S3.	2.16e+01
26	51	68.0	564	14	072736	A54R PROTEIN.	2.16e+01
27	51	68.0	581	1	058642	581AA LONG HYPOTHETICA	2.16e+01
28	51	68.0	648	6	029094	PROTEIN S (FRAGMENT).	2.16e+01
29	50	66.7	132	10	024651	CYTOTOXIC B5 (FRAGMEN	3.27e+01
30	50	66.7	345	5	045529	Y102A5B.1 PROTEIN.	3.27e+01
31	50	66.7	350	5	019556	F18E2.4 PROTEIN.	3.27e+01
32	50	66.7	398	5	090822	F09F3.10 PROTEIN.	3.27e+01
33	50	66.7	400	5	061205	T22D1.8 PROTEIN.	3.27e+01
34	50	66.7	412	2	054066	ICAA.	3.27e+01
35	50	66.7	440	2	077814	PROLIDASE (EC 3.4.13.9	3.27e+01
36	50	66.7	575	1	028180	ACETOLACTATE SYNTHASE.	3.27e+01
37	50	66.7	641	11	035646	CALPAIN 6 (CALPAIN-LIK	3.27e+01
38	50	66.7	641	11	088501	CALPAIN-LIKE PROTEASE.	3.27e+01
39	49	65.3	71	8	009381	CYTOTOXIC OXIDASE I (4.94e+01
40	49	65.3	141	5	001816	SIMILAR TO CYTOCHROME	4.94e+01
41	49	65.3	209	14	085959	GP120 (FRAGMENT).	4.94e+01
42	49	65.3	252	14	089687	VIF PROTEIN.	4.94e+01
43	49	65.3	471	5	062133	F02H6.5 PROTEIN.	4.94e+01
44	49	65.3	555	5	017388	F08F1.7 PROTEIN.	4.94e+01
45	49	65.3	974	5	023463	THE ABOVE GENBANK ENTR	4.94e+01

ALIGNMENTS

RESULT 1
ID 075326 PRELIMINARY; PRT; 666 AA.
AC 075326;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN L.
GN SEMAL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
CC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030698; G3523115; -.
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 75; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 2.45e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 152 CWNLVNGTV 160
Qy 120 CWNLVNGTV 128

RESULT 2
ID 088371 PRELIMINARY; PRT; 393 AA.
AC 088371;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN L (FRAGMENT).
GN SEMAL.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
CC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389619.

RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
viruses";
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030699; G3523117; -.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;

Query Match 86.7%; Score 65; DB 11; Length 393;
Best Local Similarity 77.8%; Pred. No. 3.70e-02; Mismatches 1; Indels 0; Gaps 0;
Matches 7; Conservative

Db 149 CWNLVNDV 157
|||:|:|
QY 120 CWNLVNGTV 128

RESULT 3
ID Q33738 PRELIMINARY; PRT; 152 AA.
AC Q33738;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE NTP PYROPHOSPHOHYDROLASE.
GN PUR7.
OS STREPTOMYCES ALBONIGER.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC12461;
RX MEDLINE; 96139493.
RA TERCERO J.A., ESPINOSA J.C., LACALLE R.A., JIMENEZ A.;
RT "The biosynthetic pathway of the aminonucleoside antibiotic
puromycin, as deduced from the molecular analysis of the pur cluster
of Streptomyces alboniger";
RT J. BIOL. CHEM. 271:1579-1590(1996).
RL EMBL; X92429; G1045226; -.
DR PROSITE; PS00893; MUTT; 1.
DR PFAM; PF00293; mutt; 1.
KW HYDROLASE.
SQ SEQUENCE 152 AA; 16820 MW; 39315A38 CRC32;

Query Match 77.3%; Score 58; DB 2; Length 152;
Best Local Similarity 55.6%; Pred. No. 9.99e-01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 37 CWDVVGAV 45
||:|:|
QY 120 CWNLVNGTV 128

RESULT 4
ID O69261 PRELIMINARY; PRT; 367 AA.
AC O69261;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE ENDO-1,4-BETA-XYLANASE (EC 3.2.1.8).
GN XYNA.
OS BACILLUS SP.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D3;
RA BRETTON C., CUMMINGS N.J., CONNERTON I.F.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Y16849; E1283264; -.
KW XYLAN DEGRADATION; HYDROLASE; GLYCOSIDASE.
SQ SEQUENCE 367 AA; 42213 MW; FD49AA71 CRC32;

Query Match 73.3%; Score 55; DB 2; Length 367;

Best Local Similarity 55.6%; Pred. No. 3.84e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 159 CWDVVNEAV 167
||:|:|:|
QY 120 CWNLVNGTV 128

RESULT 5
ID O45530 PRELIMINARY; PRT; 308 AA.
AC O45530;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE F49A5.4 PROTEIN.
GN F49A5.4
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA MORTIMORE B.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIES J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RL NATURE 368:32-38(1994).
DR EMBL; Z81542; E1346961; -.
SQ SEQUENCE 308 AA; 34893 MW; 0B06C41F CRC32;

Query Match 72.0%; Score 54; DB 5; Length 308;
Best Local Similarity 71.4%; Pred. No. 5.96e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 63 CWKLVGT 69
||:|:|
QY 120 CWNLVNGTV 126

RESULT 6
ID O16514 PRELIMINARY; PRT; 539 AA.
AC O16514;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE T05H4.9 PROTEIN.
GN T05H4.9
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BLANCHARD M.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF016452; G2315517; -;
SQ SEQUENCE 539 AA; 62652 MW; BE463856 CRC32;

Query Match 72.0%; Score 54; DB 5; Length 539;
Best Local Similarity 55.6%; Pred. No. 5.96e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 37 CWNINQAV 45
|||:|:|
QY 120 CWNLVNGTV 128

RESULT 7
ID O48845 PRELIMINARY; PRT; 134 AA.
AC O48845;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PUTATIVE CYTOCHROME B5.
GN F24L7.14.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA: VIRIDIPANTAE: STREPTOPHYTA: EMBRYOPHYTA: TRACHEOPHYTA;
OC EUPHYLLOPHYTES: SPERMATOPHYTA: MAGNOLIOPHYTA: EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-CV, COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AC003974; G2914701; -;
DR PROSITE; PS00191; CYTOCHROME_B5; 1.
KW HEME.
SQ SEQUENCE 134 AA; 15016 MW; 3EAC1E72 CRC32;

Query Match 70.7%; Score 53; DB 10; Length 134;
Best Local Similarity 55.6%; Pred. No. 9.20e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 22 CWIVNGKV 30
||:|:|
QY 120 CWNLVNGTV 128

RESULT 8
ID Q90807 PRELIMINARY; PRT; 312 AA.
AC Q90807;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE OLFATORY RECEPTOR 3 (FRAGMENT).
GN COR3.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: ARCHOSAURIA: AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE: 96184947.
RA LEBOVICI M., LAPOINTE F., ALETTA P., AYER-LE LIEVRE C.;
RT "Avian olfactory receptors: differentiation of olfactory neurons
RT under normal and experimental conditions.";
RL DEV. BIOL. 175:118-131(1996).
DR EMBL: X94743; E220348; -;
DR PFAM: PF00001; 7tm_1; 1.
FT NON_TER 312 312
SQ SEQUENCE 312 AA; 35258 MW; 6B874970 CRC32;

Query Match 70.7%; Score 53; DB 13; Length 312;
Best Local Similarity 71.4%; Pred. No. 9.20e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 141 CWRLVKG 147
|||:|:|
QY 120 CWNLVNG 126

RESULT 9
ID O83429 PRELIMINARY; PRT; 451 AA.
AC O83429;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE D-ALANINE GLYCINE PERMEASE (DAGA).
GN TP0414.
OS TREPONEMA PALLIDUM.
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; TREPONEMA.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 98332770.
RA FRASER C.M., NORRIS M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA DODSON R., GWINN M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA SODERGREN E., HARDHAM J.M., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA KHALAK H., RICHARDSON D., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RT "Complete Genome Sequence of Treponema pallidum, the Syphilis
RT Spirochete.";
RL SCIENCE 281:375-388(1998).
RN [2]
RN SEQUENCE FROM N.A.
RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF001219; G3322698; -;
SQ SEQUENCE 451 AA; 47349 MW; FOA9BA35 CRC32;

Query Match 70.7%; Score 53; DB 2; Length 451;
Best Local Similarity 50.0%; Pred. No. 9.20e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 243 CWDIVSGA 250
||:|:|:|
QY 120 CWNLVNGT 127

RESULT 10
ID Q32315 PRELIMINARY; PRT; 475 AA.
AC Q32315;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 53.7 KD PROTEIN.
OS BACILLUS THURINGIENSIS.

OG PLASMID PG13.
 OC BACTERIA: FIRMICUTES: BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 RN BACILLUS.

RP [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97405895.
 RA HOFACK L., SEURINCK J., MAHILLON J.:
 RT "Nucleotide sequence and characterization of the cryptic Bacillus
 RT thuringiensis plasmid pg13 reveal a new family of rolling circle
 RT replicons".
 RL J. BACTERIOL. 179:5000-5008(1997).
 DR EMBL: F11173; E311330; -.
 KW HYPOTHETICAL PROTEIN: PLASMID.
 SQ SEQUENCE 475 AA; 53733 MW; FCE7ABC8 CRC32;

Query Match 70.7%; Score 53; DB 2; Length 475;
 Best Local Similarity 71.4%; Pred. No. 9.20e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 466 CWHVNG 472
 ||:||||
 QY 120 CWNLVNG 126

RESULT 11
 ID O04354 PRELIMINARY; PRT; 132 AA.
 AC O04354;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE CYTOCHROME B5.
 OS BORAGO OFFICINALIS (BOURRACHE).
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
 OC ASTERIDAE; SOLANACEAE; SOLANALES; BORAGINACEAE; BORAGO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97268723.
 RA SAYANOVA O., SMITH M.A., LAPINSKAS P.A., STOBART K., DOBSON G.,
 RA CHRISTIE W.W., SHERWY P.R., NAPIER J.A.;
 RT "Expression of a borage desaturase cDNA containing an N-terminal
 RT cytochrome b5 domain results in the accumulation of high levels of
 RT delta6-desaturated fatty acids in transgenic tobacco".
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:4211-4216(1997).
 DR EMBL: U79011; G2062405; -.
 DR PROSITE: PS00191; CYTOCHROME_B5; 1.
 DR PFAM: PF00173; heme_1; 1.
 DR MENDEL: 15843; BORof; 1218; mn15843.
 KW HEME.
 SQ SEQUENCE 132 AA; 14556 MW; 3AC8F513 CRC32;

Query Match 69.3%; Score 52; DB 10; Length 132;
 Best Local Similarity 55.8%; Pred. No. 1.41e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 19 CWLLINGKV 27
 ||:||||
 QY 120 CWNLVNGTV 128

RESULT 12
 ID O45823 PRELIMINARY; PRT; 317 AA.
 AC O45823;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE T25E12.9 PROTEIN.
 GN T25E12.9.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.

RA MATTHEWS L.;
 RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.

RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
 RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., FULTON L.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans".
 RL NATURE 368:32-38(1994).
 DR EMBL: Z82052; E1349869; -.
 SQ SEQUENCE 317 AA; 35436 MW; 8BAC22E5 CRC32;

Query Match 69.3%; Score 52; DB 5; Length 317;
 Best Local Similarity 57.1%; Pred. No. 1.41e+01;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 68 CWKLITG 74
 ||:|:
 QY 120 CWNLVNG 126

RESULT 13
 ID O82111 PRELIMINARY; PRT; 513 AA.
 AC O82111;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 58.7 KD PROTEIN.
 GN F1715.10.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALE; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA VITALE D., LIGUORI R., ARGIROU A., DE SIMONE V., HOHEISEL J.,
 RA JESSE T., HEIJNEN L., VOS P., MENES H.W., MAYER K.F.X., SCHUELLER C.,
 RA BEVAN M.;
 RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT;
 RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AL031032; E1310383; -.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 513 AA; 58658 MW; 9586CC6E CRC32;

Query Match 69.3%; Score 52; DB 10; Length 513;
 Best Local Similarity 44.4%; Pred. No. 1.41e+01;
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 89 CWTLLKGGI 97
 ||:|:
 QY 120 CWNLVNGTV 128

RESULT 14
 ID O81897 PRELIMINARY; PRT; 536 AA.
 AC O81897;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE BETA-X-LAN ENDOHYDROLASE -LIKE PROTEIN.
 GN T16L1.300.

OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA: VIRIDIPANTAE: STREPTOPHYTA: EMBRYOPHYTA: TRACHEOPHYTA:
 OC EUPHYLLOPHYTES: SPERMATOPHYTA: MAGNOLIOPHYTA: EUDICOTYLEDONS: ROSIDAE;
 OC CAPPARALES: BRASSICACEAE; ARABIDOPSIS.

RN [1]
 RP SEQUENCE FROM N.A.
 RA OBERMAIER B., DEUTSCHENBAUR S., PIRAVANDI E., HOEISEL J., JESSE T.,
 RA HEIJNEN L., VOS P., MEWES H.W., MAYER K.F.X., SCHUELLER C., BEVAN M.;
 RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AL031394; E1318072; -
 KW HYDROLASE.
 SQ SEQUENCE 536 AA; 61259 MW; 4F8CD66 CRC32;

Query Match 69.3%; Score 52; DB 10; Length 536;
 Best Local Similarity 44.4%; Pred. No. 1.41e+01;
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 89 CWTLLKGGI 97
 |||:::
 QY 120 CWNLVNGTV 128

RESULT 15

ID Q99129 PRELIMINARY; PRT; 1150 AA.
 AC Q99129;
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
 DE SMUT FUNGUS (CLONE PM1-1) MYPI (CLONE PM1-1).
 GN MYPI.
 OS USTILAGO MAYDIS (SMUT FUNGUS).
 OC EUKARYOTA; FUNGI; BASIDIOMYCOTA; USTILAGINOMYCETES; USTILAGO.
 OC USTILAGINOMICETIDAE; USTILAGINALES; USTILAGINACEAE; USTILAGO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=518;
 RX MEDLINE; 96109597.
 RA GIASSON L., KRONSTAD J.W.;
 RT "Mutations in the mypi gene of Ustilago maydis attenuate mycelial
 growth and virulence."
 RL GENETICS 141:491-501(1995).
 DR EMBL: L33919; G886416; -
 SQ SEQUENCE 1150 AA; 123387 MW; 68647042 CRC32;

Query Match 69.3%; Score 52; DB 3; Length 1150;
 Best Local Similarity 66.7%; Pred. No. 1.41e+01;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 822 CWPLVRVTV 830
 |||||
 QY 120 CWNLVNGTV 128

Search completed: Thu Jul 8 18:23:17 1999
 Job time : 16 secs.

THIS PAGE BLANK (USPTO)

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:26:18 1999; MasPar time 9.45 Seconds
Tabular output not generated. 17.997 Million cell updates/sec

Title: >US-09-041-236-2
Description: (175-182) from US09041236.pep (10 of 45)
Perfect Score: 57

Sequence: 1 RGESELYT 8

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 16.291; Variance 41.167; scale 0.396

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	47	82.5	455	17	Bombyx mori mature LP	6.17e+01
2	47	82.5	467	17	Bombyx mori full leng	6.17e+01
3	46	80.7	20	7	Platelet adhesion inh	8.22e+01
4	44	77.2	356	36	Human glycosaminoglyc	1.45e+02
5	43	75.4	193	11	P. aeruginosa algu.	1.91e+02
6	43	75.4	193	11	P. aeruginosa algu.	1.91e+02
7	43	75.4	351	19	Toluene ortho-monooxy	1.91e+02
8	43	75.4	356	36	Glycosaminoglycan sul	1.91e+02
9	43	75.4	697	1	Protein kinase C-IV (1.91e+02
10	42	73.7	110	7	Tryptophan aporepress	2.52e+02
11	42	73.7	226	22	H. pylori secreted or	2.52e+02
12	42	73.7	236	14	V. fischeri flavin red	2.52e+02
13	42	73.7	339	22	H. pylori secreted or	2.52e+02
14	42	73.7	1044	29	Mammalian novel class	2.52e+02
15	42	73.7	1044	33	Human phosphatidylino	2.52e+02
16	42	73.7	1284	1	Sequence encoded by a	2.52e+02

17	41	71.9	193	23	W14145	Erythropoietin varian	3.32e+02
18	41	71.9	200	16	R82900	Mouse B7-1 (Igv-like	3.32e+02
19	41	71.9	214	16	R82901	Mouse B7-2 antigen.	3.32e+02
20	41	71.9	306	39	W73641	Mouse B7-2 antigen.	3.32e+02
21	41	71.9	306	16	R82893	Mouse B7-1 alternati	3.32e+02
22	41	71.9	306	13	R67990	Murine B lymphocyte a	3.32e+02
23	41	71.9	320	16	R82892	Mouse B7-1 alternati	3.32e+02
24	41	71.9	456	28	W25764	Xenopus DNA replicati	3.32e+02
25	40	70.2	36	6	R33060	Peptide from CD3 epsi	4.35e+02
26	40	70.2	54	6	R33052	Peptide from CD3 epsi	4.35e+02
27	40	70.2	55	6	R33053	Peptide from CD3 epsi	4.35e+02
28	40	70.2	55	6	R33066	Peptide from CD3 epsi	4.35e+02
29	40	70.2	55	6	R33065	Peptide from CD3 epsi	4.35e+02
30	40	70.2	113	35	W77045	Rat aortic-preferen	4.35e+02
31	40	70.2	113	35	W77047	Human aortic-preferen	4.35e+02
32	40	70.2	113	23	W10566	Human aortic-preferen	4.35e+02
33	40	70.2	117	8	R38777	Active site peptide f	4.35e+02
34	40	70.2	117	5	R25524	Tobacco Ring spot Vir	4.35e+02
35	40	70.2	178	34	W29658	Homo sapiens BH272_3	4.35e+02
36	40	70.2	272	6	R29883	HCV NS4-NS5 peptide N	4.35e+02
37	40	70.2	572	8	R40843	Bilirubin oxidase.	4.35e+02
38	40	70.2	601	35	W77049	Mouse striated muscle	4.35e+02
39	40	70.2	863	6	R29881	HCV NS4-NS5 peptide 2	4.35e+02
40	40	70.2	1693	35	W76368	Hepatitis E virus hol	4.35e+02
41	40	70.2	1693	37	W81519	Hepatitis E virus (HE	4.35e+02
42	40	70.2	1693	34	W71209	Protein encoded by OR	4.35e+02
43	40	70.2	1693	36	W80196	Protein encoded by OR	4.35e+02
44	40	70.2	1693	19	R91813	Hepatitis E virus str	4.35e+02
45	40	70.2	2510	6	R29527	HCV antigen T7N1-30.	4.35e+02

ALIGNMENTS

RESULT 1

ID R89136 standard; protein; 455 AA.
AC R89136;
DT 22-AUG-1996 (first entry)
DE Bombyx mori mature LPS-binding protein.
KW lipopolysaccharide binding protein; Bombyx mori; haemolymph; primer;
KW Enterobacter cloacae; PCR; polymerase chain reaction; amplification;
KW probe; LPS-complex; septic shock; injection; transgenic plant; vine;
KW tobacco; tomato; potato; fungal infection; fungus.
OS Bombyx mori.
EH key. Location/Qualifiers
FT modified_site 170
FT /note= "putative N-glycosylation site"
PN FR2721032-A1.
PD 15-DEC-1995.
PF 09-JUN-1994; 007083
PR 09-JUN-1994; FR-007083.
PA (INSP) INST PASTEUR.
PI Brey PT, Lee W;
DR WPI; 96-060094/07.
PT New protein from Bombyx mori that binds bacterial
PT lipopolysaccharide - esp. used to treat septic shock, also DNA
PT encoding it, for producing transgenic plant(s) resistant to fungal
PT attack
PS Claim 4; Page 41-42; 53pp; French.
CC This is the amino acid sequence of a mature lipopolysaccharide (LPS)
CC binding protein from Bombyx mori. The protein was isolated from the
CC haemolymph of fifth stage B.mori larvae after injection with Enterobacter
CC cloacae strain 57-9. Partial amino acid sequence was used to generate
CC PCR primers (T10280-1). These amplified a fragment of the gene used as a
CC probe to obtain the full length gene (T10879) by screening a cDNA library
CC derived from B.mori previously injected with heat-killed E.coli. The
CC protein or fragments of it, can be used to treat diseases associated with
CC LPS complexes e.g. septic shock, to remove LPS from products used for
CC injection and to protect transgenic plants e.g. vines, tobacco, tomato or
CC potato, against fungal infections.
SQ Sequence 455 AA;

Query Match 82.5%; Score 47; DB 17; Length 455;
Best Local Similarity 62.5%; Pred. No. 6.17e+01;

```

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 287 rgnaclys 294
QY 175 RGESELYT 182

RESULT 2
ID R59137 standard; Protein; 467 AA.
AC R59137;
DE Bombyx mori full length LPS-binding protein.
KW lipopolysaccharide binding protein; Bombyx mori; haemolymph; primer;
KW Enterobacter cloacae; PCR; polymerase chain reaction; amplification;
KW probe; LPS-complex; septic shock; injection; transgenic plant; vine;
KW tobacco; tomato; potato; fungal infection; fungus.
OS Bombyx mori.
FH Key
FT peptide 1..12
FT 13..467
FT peptide
FT modified_site 182
FT /note= "putative N-glycosylation site"
FT FR2721032-A1.
FT 15-DEC-1995.
FT 09-JUN-1994; 007083.
FT 09-JUN-1994; FR-007083.
FT (INSP ) INST PASTEUR.
FT PI Brey PT, Lee W;
FT WPI: 96-060094/07.
FT N-PSDB: T10279.
FT New protein from Bombyx mori that binds bacterial
FT lipopoly-saccharide - esp. used to treat septic shock, also DNA
FT encoding it, for producing transgenic plant(s) resistant to fungal
FT attack
FT Claim 5; Page 42-43; 53pp; French.
FT This is the amino acid sequence of the full length lipopolysaccharide
FT (LPS) binding protein from Bombyx mori. The protein was isolated from the
FT hemolymph of fifth stage B.mori larvae after infection with Enterobacter
FT cloacae strain 57-9. Partial amino acid sequence was used to generate
FT PCR primers (T10280-1). These amplified a fragment of the gene used as
FT a probe to obtain the full length gene by screening a cDNA library
FT derived from B.mori previously injected with heat-killed E.cloacae. The
FT protein or fragments of it, can be used to treat diseases associated with
FT LPS complexes e.g. septic shock, to remove LPS from products used for
FT injection and to protect transgenic plants e.g. vines, tobacco, tomato or
FT potato, against fungal infections.
FT Sequence 467 AA;

Query Match 82.5%; Score 47; DB 17; Length 467;
Best Local Similarity 62.5%; Pred. No. 6.17e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 299 rgnaclys 306
QY 175 RGESELYT 182

RESULT 3
ID R35065 standard; peptide; 20 AA.
AC R35065;
DE 04-AUG-1993 (first entry)
DE Platelet adhesion inhibitor fusion peptide #1.
KW Cell; aggregation; adhesion; platelets; disease; platelet dependant;
KW thrombosis; restenosis; spacer.
OS Synthetic.
FH Key
FT peptide 3..4
FT /note= "Spacer peptide"
FT W09307169-A.
FT 15-APR-1993.
FT 05-OCT-1992; U08481.

PR 04-OCT-1991; US-771768.
PA (CHIR ) CHIRON CORP.
PI Devlin JJ, Doyle M, Fong S;
DR WPI: 93-134382/16.
PT Peptide inhibitors of platelet adhesion - for prevention and
PT treatment of thrombosis and restenosis
PS Claim 3; Page 46; 67pp; English.
CC The sequences given in R35065-66 are fusion peptides which interfere
CC with cell aggregation or adhesion, pref. of platelets. They are
CC useful for treating or preventing diseases selected from platelet
CC dependent thrombosis and restenosis.
CC Sequence 20 AA;

Query Match 80.7%; Score 46; DB 7; Length 20;
Best Local Similarity 50.0%; Pred. No. 8.22e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 1 rgdalifs 8
QY 175 RGESELYT 182

RESULT 4
ID W72573 standard; Protein; 356 AA.
AC W72573;
DT 06-JAN-1999 (first entry)
DE Human glycosaminoglycan sulphate group transferase.
KW Glycosaminoglycan sulphate group transferase; chinese hamster;
KW L-iduronic acid residue; sulphate group receptor glycosaminoglycan;
KW heparan sulphate 2-O-sulphate group transferase; HS2ST.
OS Homo sapiens.
KW J10257896-A.
PD 29-SEP-1998.
PF 04-JUN-1997; 146815.
PR 17-JAN-1997; JP-006522.
PA (SEKG ) SEIKAGAKU KOGYO CO LTD.
DR WPI: 98-575907/49.
DR N-PSDB: V68827.
PT A polynucleotide encoding glycosaminoglycan sulphate group
PT transferase - useful for the recombinant production of the enzyme
PS Claim 5; Page 18-19; 22pp; Japanese.
CC The present sequence represents a specifically claimed human
CC glycosaminoglycan sulphate group transferase. The present invention
CC describes a DNA molecule coding at least part of a polypeptide of
CC glycosaminoglycan sulphate group transferase having the 356 amino acid
CC sequence as shown in W72571 to W72573, and optionally having a
CC replacement, deletion, or insertion of at least one amino acid (aa)
CC residue but still retaining the enzymic activity of transferring a
CC sulphate group from a sulphate group donor to the 2-OH of a L-iduronic
CC acid residue contained in a sulphate group receptor glycosaminoglycan.
CC The nucleic acid can be used for the recombinant production of the
CC enzyme, especially for the production of heparan sulphate 2-O-sulphate
CC group transferase (HS2ST).
CC Sequence 356 AA;

Query Match 77.2%; Score 44; DB 36; Length 356;
Best Local Similarity 71.4%; Pred. No. 1.45e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 273 rgately 279
QY 175 RGESELYT 181

RESULT 5
ID R57060 standard; Protein; 193 AA.
AC R57060;
DT 21-MAR-1995 (first entry)
DE P. aeruginosa alga.
KW alga; mucoidy; cystic fibrosis; diagnostic.
OS Pseudomonas aeruginosa.
FN W09418223-A.
PD 18-AUG-1994.

```


PF 14-FEB-1994; U02034.
 PR 12-FEB-1993; US-017114.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Deretic V, Martin DW;
 DR WPI; 94-279677/34.
 PT New mucoid-related Pseudomonas aeruginosa genes - used to develop
 PT prods for detecting conversion to mucoidy in P. aeruginosa
 PT infecting cystic fibrosis patients
 PS Disclosure; Page 121-122; 147pp; English.
 CC Genetic studies were used to characterize genes within the
 CC chromosomal region at 67.5 min which are critical in determining the
 CC mucoid status of P. aeruginosa. 3 tightly linked genes, algU, muca
 CC and mucB, were identified and demonstrated to participate in the
 CC control of mucoidy.
 SQ Sequence 193 AA;

 Query Match 75.4%; Score 43; DB 11; Length 193;
 Best Local Similarity 62.5%; Pred. No. 1.91e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 Db 66 rgdsafyt 73
 ||:|:|
 Qy 175 RGESELYT 182

 RESULT 6
 ID R57059 standard; Protein: 193 AA.
 AC R57059;
 DT 21-MAR-1995 (first entry)
 DE P. aeruginosa algU.
 KW algU; mucoidy; cystic fibrosis; diagnostic.
 OS Pseudomonas aeruginosa.
 PN W09418223-A.
 PD 18-AUG-1994.
 PF 14-FEB-1994; U02034.
 PR 12-FEB-1993; US-017114.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Deretic V, Martin DW;
 DR WPI; 94-279677/34.
 DR N-PSDB; Q70242.
 PT New mucoid-related Pseudomonas aeruginosa genes - used to develop
 PT prods for detecting conversion to mucoidy in P. aeruginosa
 PT infecting cystic fibrosis patients
 PS Disclosure; Fig.1; 147pp; English.
 CC Genetic studies were used to characterize genes within the
 CC chromosomal region at 67.5 min which are critical in determining the
 CC mucoid status of P. aeruginosa. 3 tightly linked genes, algU, muca
 CC and mucB, were identified and demonstrated to participate in the
 CC control of mucoidy.
 SQ Sequence 193 AA;

 Query Match 75.4%; Score 43; DB 11; Length 193;
 Best Local Similarity 62.5%; Pred. No. 1.91e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 Db 66 rgdsafyt 73
 ||:|:|
 Qy 175 RGESELYT 182

 RESULT 7
 ID W05804 standard; Protein: 351 AA.
 AC W05804;
 DT 29-JAN-1997 (first entry)
 DE Toluene ortho-monoxygenase subunit tomA5.
 KW tom; pTOM; self-transmissible; constitutive; bioreactor; pollutant;
 KW breakdown; trichloroethylene; TCE; degradation.
 OS Pseudomonas cepacia strain PRI-23.
 PN US5543317-A.
 PD 06-AUG-1996.
 PF 02-MAY-1991; 694718.
 PR 02-MAY-1991; US-694718.
 PR 15-DEC-1993; US-167457.

PR 06-OCT-1994; US-319387.
 PA (FRAN/) FRANCESCONI S C.
 PA (SHIE/) SHIELDS M S.
 PI Francesconi SC, Shields MS;
 DR WPI; 96-370640/37.
 DR N-PSDB; T44457.
 PT Microorganisms transformed with P. cepacia PRI-23 Tom enzyme gene -
 PT are useful for degradation of chloroaliphatic cpds. and aromatics
 PS Claim 1; Column 33-36; 25pp; English.
 CC The present sequence is that of toluene ortho-monoxygenase subunit
 CC tomA5 encoded by T44457, isolated from Pseudomonas cepacia strain
 CC PRI-23. The Tom gene is present on a large self-transmissible plasmid
 CC denoted pTOM. The enzyme is capable of degrading trichloroethylene
 CC (TCE), a hazardous pollutant. The pTOM plasmid is transmissible and
 CC expressable in other bacteria, thus many bacteria can be genetically
 CC altered to constitutively degrade TCE, esp. in bioreactors or
 CC TCE-contaminated environments. P. cepacia PRI-23 contg. pTOM does not
 CC need exogenous chemical inducers and is capable of functioning under a
 CC diverse set of conditions. Also it does not require an inducer that is a
 CC co-substrate for the enzyme required to break down TCE, in effect, it is
 CC not subject to competitive inhibition.
 SQ Sequence 351 AA;

 Query Match 75.4%; Score 43; DB 19; Length 351;
 Best Local Similarity 71.4%; Pred. No. 1.91e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 Db 244 rneaely 250
 |:|:|
 Qy 175 RGESELY 181

 RESULT 8
 ID W72572 standard; Protein: 356 AA.
 AC W72572;
 DT 06-JAN-1999 (first entry)
 DE Glycosaminoglycan sulphate group transferase; chinese hamster;
 KW Glycosaminoglycan sulphate group transferase; chinese hamster;
 KW L-Iduronic acid residue; sulphate group receptor glycosaminoglycan;
 KW heparan sulphate 2-O-sulphate group transferase; HS2ST.
 OS Crictetus sp.
 PN J10257896-A.
 PD 29-SEP-1998.
 PF 04-JUN-1997; 146815.
 PR 17-JAN-1997; JP-006522.
 PA (SEBK) SEIKAGAKU KOGYO CO LTD.
 DR WPI; 98-575907/49.
 DR N-PSDB; V66826.
 PT A polynucleotide encoding glycosaminoglycan sulphate group
 PT transferase - useful for the recombinant production of the enzyme
 PS Claim 3; Page 15-16; 22pp; Japanese.
 CC The present sequence represents a specifically claimed glycosaminoglycan
 CC sulphate group transferase from Chinese hamster. The present invention
 CC describes a DNA molecule coding at least part of a polypeptide of
 CC glycosaminoglycan sulphate group transferase having the 356 amino acid
 CC sequence as shown in W72571 to W72573, and optionally having a
 CC replacement, deletion, or insertion of at least one amino acid (aa)
 CC residue but still retaining the enzymic activity of transferring a
 CC sulphate group from a sulphate group donor to the 2-OH of a L-iduronic
 CC acid residue contained in a sulphate group receptor glycosaminoglycan.
 CC The nucleic acid can be used for the recombinant production of the
 CC enzyme, especially for the production of heparan sulphate 2-O-sulphate
 CC group transferase (HS2ST).
 SQ Sequence 356 AA;

 Query Match 75.4%; Score 43; DB 36; Length 356;
 Best Local Similarity 57.1%; Pred. No. 1.91e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 Db 273 rgatdly 279
 |:|:|
 Qy 175 RGESELY 181

```

RESULT 9
ID P82018 standard; protein; 697 AA.
AC P82018;
DT 16-OCT-1990 (first entry)
DE Protein kinase C-IV (PKC-IV)
KW Protein kinase C; PKC; cancer; ds.
OS Rattus sp.
PN W08801303-A.
PD 25-FEB-1988.
PF 13-AUG-1987; 002005.
PR 13-AUG-1986; US-896476.
PA (GENE-) Genetics Inst Inc.
PI Knopf JL;
DR WPI: 88-064018/09.
DR N-NSPD: N80336.
PT New DNA sequences coding for protein C enzyme -
PT and new expressed polypeptide(s), useful for detecting tumour
PT promoting activity of test cpds.
PS Claim 1; Fig 2; 36pp; English.
CC PKC product can be used in determining tumour promoting properties
CC of a test sample, and probes derived from the sequence may be used
CC in isolating polypeptides from other species.
SQ Sequence 697 AA;

Query Match 75.4%; Score 43; DB 1; Length 697;
Best Local Similarity 62.5%; Pred. No. 1.91e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 371 rgsdelya 378
|| |||
QY 175 RGESELYT 182

RESULT 10
ID R35533 standard; protein; 110 AA.
AC R35533;
DT 09-JUL-1993 (first entry)
DE Tyroptophan aporepressor (3.13).
KW TA: N-terminal; C-terminal; complement C3; ligand; binding site;
KW vaccine; hybrid; cell receptor.
OS Synthetic.
FH Key Location/Qualifiers
FT region 1..66 /note= "TA N-terminal portion"
FT region 67..75 /note= "complement C3 ligand binding site"
FT region 76..110 /note= "TA C-terminal"
FT W09300356-A.
PN W09300356-A.
PD 07-JAN-1993.
PF 19-JUN-1992; U05224.
PR 21-JUN-1991; US-720222.
PA (CALB-) CALIFORNIA INST BIOLOGICAL RES.
PI Bourdon M, Lernerhardt W, Yoderian P;
DR WPI: 93-036327/04.
PT Tyroptophan aporepressor used in vaccine prodn., etc. - contains
PT within its' amino acid sequence, a peptide segment heterologous
PT to the aporepressor located on aq. solvent-accessible surface
PS Claim 5 + disclosure; Page 69 + 23; 82pp; English.
CC The hybrid protein contg. receptor binding sites is active as ligand
CC for mammalian cell receptors and can be used for a variety of
CC applications including treatment of diseases resulting from receptor/
CC ligand dysfunction, and in vaccine prodn.
SQ Sequence 110 AA;

Query Match 73.7%; Score 42; DB 7; Length 110;
Best Local Similarity 71.4%; Pred. No. 2.52e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 63 rgenqly 69
||| |||
QY 175 RGESELY 181

RESULT 11
ID W20422 standard; Protein; 236 AA.
AC W20422;
DT 14-JUL-1997 (first entry)
DE H. pylori secreted or periplasmic protein 33595708.aa.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT misc_difference 47 /label= unknown
FT /note= "encoded by GRA"
FT misc_difference 183 /label= unknown
FT /note= "encoded by YTT"
FT misc_difference 210 /label= unknown
FT /note= "encoded by GAY"
FT misc_difference 224 /label= unknown
FT /note= "encoded by TYA"
FT W09640893-A1.
PN W09640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI: 97-052306/05.
DR N-PSDB: T67598.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 72; Page 601; 1481pp; English.
CC The present sequence encodes a Helicobacter pylori secreted or
CC periplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 226 AA;

Query Match 73.7%; Score 42; DB 22; Length 226;
Best Local Similarity 57.1%; Pred. No. 2.52e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 97 rgdfldly 103
||| |||
QY 175 RGESELY 181

RESULT 12
ID R70423 standard; Protein; 236 AA.
AC R70423;
DT 13-DEC-1995 (first entry)
DE V.fischeri flavin reductase specific protein sequence.
KW V.fischeri reductase; Vibrio fischeri; primer; PCR; amplification; probe;
KW Xenorhabdus luminescens; expression plasmid; FMN reducing activity;
KW E.coli.
OS Vibrio fischeri ATCC-7744.
FH Key Location/Qualifiers

```

ET misc_difference 149 /note= "encoded by CAA; in R74669 there is an Arg
FT residue at this position"
FT
PN J07079783-A.
PD 28-MAR-1995.
PF 28-JUN-1993; 181850.
PA (CHCC) CHISSO CORP.
DR WPI; 95-157853/21.
DR N-PSDB; Q88283.
PT Flavin reductase gene from *Vibrio fischeri* - useful for production
PT of enzyme with FMN reducing activity
PS Claim 2: Page 6-7; 10pp; Japanese.
CC The specific amino acid sequence of the flavin reductase (fre) protein
CC from *Vibrio fischeri* strain ATCC-7744. The gene was isolated from
CC genomic DNA extracted from *V. fischeri* using primers T01084-5 based on
CC sequence from the *Xenorhabdus luminescens* (ATCC-29999) fre gene. The
CC resultant fragment was cloned into pUC8 and used as a probe to obtain the
CC full length gene sequence (Q88284) of the *V. fischeri* fre gene from a
CC lambda phage library. The gene was inserted into the expression plasmid
CC pUC13 and transformed into *E. coli* D1210. The resultant transformant
CC *E. coli* D1210/pfFR1 was used for the IPTG-inducible expression of the fre
CC gene. The gene encodes a protein with FMN reducing activity.
SQ Sequence 236 AA;

Query Match 73.7%; Score 42; DB 14; Length 236;
Best Local Similarity 62.5%; Pred. No. 2.52e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 145 rdecelya 152
QY |||||:
175 RGESELY 182

RESULT 13
ID W20906 standard; protein; 339 AA.
AC W20906;
DT 18-JUL-1997 (first entry)
DE H. pylori secreted or periplasmic protein, 14eell217orf1.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.
PN W09840893-AI.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglindh OT, Smith D, Meilgaard BL;
DR WPI; 97-052306/05.
DR N-PSDB; T68159.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 72; Page 1306-1307; 1481pp; English.
CC The present sequence is a H. pylori secreted or periplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in *E. coli* hosts.
SQ Sequence 339 AA;

Query Match 73.7%; Score 42; DB 22; Length 339;

Best Local Similarity 57.1%; Pred. No. 2.52e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 178 rgdgly 184
QY ||:|:
175 RGESELY 181

RESULT 14
ID W46625 standard; protein; 1044 AA.
AC W46625;
DT 22-JUN-1998 (first entry)
DT 22-JUN-1998 (revised entry)
DE Mammalian novel class I PI3 kinase designated p110-delta.
KW Autophosphorylate; PI3; phosphatidylinositol 3-hydroxy; kinase;
KW activity; p110-delta; melanoma; lipid kinase; metastase;
KW phosphoinositide specificity; regulation; diagnosis; motility;
KW cancer cell; control.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 292..311
FT /note= "proline-rich region"
FT Domain 400..439
FT /note= "basic region leucine zipper-like domain"
PN W09746688-AI.
PD 11-DEC-1997.
PF 30-MAY-1997; G01471.
PR 01-JUN-1996; GB-011460.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Vanhasebroeck B, Waterfield MD;
DR WPI; 98-042196/04.
DR P-PSDB; V16533.
PT Auto-phosphorylating peptide with phosphatidyl:inositol
PT 3-kinase-like activity, designated p110delta - useful for
PT controlling cell motility, particularly of metastatic cancer cells
PS Disclosure; Fig 1A; 72pp; English.
CC The present sequence represents a novel autophosphorylating protein
CC that has PI3 (phosphatidylinositol 3-hydroxy) kinase activity, and
CC is designated p110-delta. The protein is expressed selectively in
CC white blood cells and melanomas. The protein is a lipid kinase with
CC broad phosphoinositide specificity and specific tissue localisation,
CC possibly involved in regulation of melanoma metastases (it is not
CC expressed in normal melanocytes). DNA fragments encoding the present
CC sequence are used to detect tissue-specific expression, and a
CC similar analysis can be done at the protein level using antibodies in
CC standard immunoassays. These assays are particularly used for diagnosing
CC and predicting motility/invasiveness of metastatic cancer cells. The
CC protein can be used in human or veterinary medicine for controlling
CC motility of cells, where the protein increases motility while antisense
CC sequences are used to reduce it.
CC note: this record is revised to include second DR line which gives the
CC nucleotide sequence.
SQ Sequence 1044 AA;

Query Match 73.7%; Score 42; DB 29; Length 1044;
Best Local Similarity 71.4%; Pred. No. 2.52e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 518 rgsgeely 524
QY ||:|:
175 RGESELY 181

RESULT 15
ID W58570 standard; Protein; 1044 AA.
AC W58570;
DT 12-OCT-1998 (first entry)
DE Human phosphatidylinositol 3-kinase p110 catalytic delta subunit.
KW Phosphatidylinositol 3-kinase; p110 delta; human; immune system;
KW carcinogenesis; diagnosis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Domain 723..1044

```

FT      Binding_site      /note= "catalytic domain"
FT      141..310
FT      /note= "proposed Ras binding site"
PN      WO9823760-A1.
PD      04-JUN-1998.
PF      25-NOV-1997; U21655.
PR      25-NOV-1996; US-777405.
PA      (ICOS-) ICOS CORP.
PI      Chantry DH, Hoekstra MF, Holtzman DA;
DR      WPI; 98-322735/28.
DR      N-PSDB; V31340.
PT      New phosphatidylinositol 3-kinase catalytic subunit - used to
PT      develop products for modulating kinase activity in immune system
PT      signalling and in carcinogenesis
PS      Claim 11: Page 33-37; 53pp; English.
CC      This is the deduced amino acid sequence of the catalytic p110
CC      delta subunit of human phosphatidylinositol 3-kinase (PI 3-kinase).
CC      The sequence was deduced from a composite cDNA clone (see V58570)
CC      derived from peripheral blood mononuclear cell and macrophage
CC      cDNA. The following are claimed: (1) a purified and isolated
CC      polynucleotide (PN) encoding p110 delta; (2) a vector comprising a
CC      DNA as in (1); (3) a host cell stably transformed or transfected
CC      with a DNA as in (1); (4) PN encoding a lipid kinase, and
CC      hybridising to PN having the 5220 bp sequence; (5) a purified and
CC      isolated p110 delta polypeptide as in (4); (6) an antibody
CC      specifically immunoreactive with p110 delta; (7) a hybridoma cell
CC      line (especially 208F (HB 12200) producing a monoclonal antibody as
CC      in (6); and (8) a humanised antibody as in (6). p110 delta has
CC      kinase activity and may play a role in PI 3-kinase mediated
CC      signalling in the immune system and in carcinogenesis. The
CC      products can be used to develop agents that modulating p110 delta
CC      kinase activity and to develop diagnostic reagents (claimed). They
CC      may also be used for detection and diagnosis of p110 delta in a
CC      biological sample.
SQ      Sequence 1044 AA;

Query Match      73.7%; Score 42; DB 33; Length 1044;
Best Local Similarity 71.4%; Pred. No. 2.52e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 518 rgsely 524
QY 175 RGESELY 181
||:|||

Search completed: Thu Jul 8 18:26:33 1999
Job time : 15 secs.

```

W P E R E L L (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:25:49 1999; Maspar time 5.26 Seconds
Tabular output not generated. 60.923 Million cell updates/sec

Title: >US-09-041-236-2
Description: (175-182) from US09041236.pep (10 of 45)
Perfect Score: 57
Sequence: 1 RGESELYT 8
Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.285; Variance 24.107; scale 0.924

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	50	87.7	250	2	A70175	1.03e+00
2	49	86.0	67	2	virB protein - versin	1.77e+00
3	47	82.5	31	2	phospholipase D - cab	5.12e+00
4	47	82.5	467	2	Gram-negative bacteri	5.12e+00
5	46	80.7	403	2	probable recF protein	8.59e+00
6	45	78.9	489	2	probable monooxygenas	1.43e+01
7	45	78.9	1343	2	DNA-directed RNA poly	1.43e+01
8	44	77.2	189	2	transcription initiat	2.36e+01
9	44	77.2	192	2	rpoE protein homolog	2.36e+01
10	44	77.2	254	2	emerin - human	2.36e+01
11	44	77.2	460	2	RAD57 protein - yeast	2.36e+01
12	44	77.2	749	2	ATP-dependent helicase	2.36e+01
13	44	77.2	863	2	gag polyprotein - hum	2.36e+01
14	44	77.2	1353	2	hypothetical protein	2.36e+01
15	44	77.2	1694	2	hypothetical 191.1K p	2.36e+01
16	43	75.4	78	2	ribosomal protein L38	3.87e+01
17	43	75.4	191	2	RNA polymerase sigma	3.87e+01
18	43	75.4	193	2	transcription initiat	3.87e+01
19	43	75.4	212	1	stringent starvation	3.87e+01
20	43	75.4	291	2	methionyl aminopeptid	3.87e+01
21	43	75.4	415	2	levansucrase - Erwini	3.87e+01
22	43	75.4	582	2	hypothetical protein	3.87e+01
23	43	75.4	602	2	hypothetical protein	3.87e+01

24	43	75.4	660	2	S54746	cyck protein - Bradyr	3.87e+01
25	43	75.4	682	1	KIBOC	protein kinase C (EC	3.87e+01
26	43	75.4	697	1	KIRTOC	protein kinase C (EC	3.87e+01
27	43	75.4	697	2	JN0548	protein kinase C (EC	3.87e+01
28	43	75.4	697	2	D24664	protein kinase C (EC	3.87e+01
29	43	75.4	697	1	KIRBOC	protein kinase C (EC	3.87e+01
30	43	75.4	711	2	F65201	phosphoenolpyruvate-p	3.87e+01
31	43	75.4	715	2	E71073	hypothetical protein	3.87e+01
32	43	75.4	857	2	E71377	probable ATP-dependen	3.87e+01
33	43	75.4	1279	2	E69681	peptide synthetase pp	3.87e+01
34	43	75.4	201	2	S29907	hypothetical protein	6.27e+01
35	42	73.7	288	2	D69144	pyruvate formate-lyas	6.27e+01
36	42	73.7	363	2	S32112	nitrite reductase (cy	6.27e+01
37	42	73.7	378	2	A64839	cytochrome oxidase (E	6.27e+01
38	42	73.7	506	2	C64377	hypothetical protein	6.27e+01
39	42	73.7	540	2	C71260	hypothetical protein	6.27e+01
40	42	73.7	553	2	B64622	osmoprotection protei	6.27e+01
41	42	73.7	553	2	G71892	probable osmoprotecti	6.27e+01
42	42	73.7	721	2	A33319	microtubule-associate	6.27e+01
43	42	73.7	2364	2	A56577	microtubule-associate	6.27e+01
44	42	73.7	2464	1	QRMSD1	hypothetical protein	6.27e+01
45	42	73.7	3051	2	S42373	hypothetical protein	6.27e+01

ALIGNMENTS

RESULT 1

ENTRY A70175 #type complete
TITLE chaperonin homolog - Lyme disease spirochete
ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease spirochete

DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998

ACCESSIONS A70175

REFERENCE A70100

#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.

#journal Nature (1997) 390:580-586

#title Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

#cross-references MUID:98065943

#accession A70175

#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA

#residues 1-250 #label KLE

##cross-references GB:AE000783; NID:g2688515; PID:g2688517;

##experimental_source strain B31

SUMMARY #length 250 #molecular-weight 29473 #checksum 6412

Query Match 87.7%; Score 50; DB 2; Length 250;

Best Local Similarity 87.5%; Pred. No. 1.03e+00;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 99 RKGSELYT 106

QY 175 RGESELYT 182

RESULT 2

ENTRY A32242 #type fragment

TITLE virB protein - Versinia enterocolitica (fragment)

ORGANISM #formal_name Versinia enterocolitica

DATE 21-May-1990 #sequence_revision 21-May-1990 #text_change 23-Mar-1993

ACCESSIONS A32242
REFERENCE A91896
#authors Cornelis, G.; Sluifers, C.; Lambert de Rouvroit, C.;
#journal J. Bacteriol. (1989) 171:254-262
#title Homology between virF, the transcriptional activator of the
Yersinia virulence regulon, and arae, the Escherichia coli
arabinose operon regulator.
#cross-references MUID:89123026
#accession A32242
#molecule_type DNA
##residues 1-67 #label COR
SUMMARY #length 67 #checksum 1700
Query Match 86.0%; Score 49; DB 2; Length 67;
Best Local Similarity 85.7%; Pred. No. 1.77e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 16 RGESELY 22
1111111
QY 175 RGESELY 181
RESULT 3
ENTRY S36119 #type complete
TITLE phospholipase D - cabbage
ORGANISM #formal_name Brassica oleracea var. capitata #common_name
cabbage
DATE 10-Dec-1993 #sequence_revision 27-Feb-1997 #text_change
13-Mar-1997
ACCESSIONS S36119
REFERENCE S36119
#authors Aboualham, A.; Riviere, M.; Teissere, M.; Verger, R.
#journal Biochim. Biophys. Acta (1993) 1158:1-7
#title Improved purification and biochemical characterization of
phospholipase D from cabbage.
#cross-references MUID:93357261
#accession S36119
##status preliminary
##molecule_type protein
##residues 1-31 #label ABO
SUMMARY #length 31 #molecular-weight 3388 #checksum 8261
Query Match 82.5%; Score 47; DB 2; Length 31;
Best Local Similarity 62.5%; Pred. No. 5.12e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 9 KGETELY 16
1111111
QY 175 RGESELY 182
RESULT 4
ENTRY JC6150 #type complete
TITLE Gram-negative bacteria-binding protein precursor - silkworm
ORGANISM #formal_name Bombyx mori #common_name silkworm
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
13-Nov-1998
ACCESSIONS JC6150; PC6039
REFERENCE JC6150
#authors Lee, W.J.; Lee, J.D.; Kravchenko, V.V.; Ulevitch, R.J.; Brey,
P.T.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:7888-7893
#title Purification and molecular cloning of an inducible
Gram-negative bacteria-binding protein from the silkworm,
Bombyx mori.
#cross-references MUID:96353914
#accession JC6150
##molecule_type mRNA
##residues 1-467 #label LEE1
##cross-references GB:L38591; NID:g1209252; PID:g1209253
#accession PC6039
##molecule_type protein

##residues 42-53;56-83;88-100 #label LEE2
##experimental_source fat body
COMMENT This is a soluble protein having affinity for the Gram-negative
bacterial cell wall. It is inducible during injury or infection
as an acute phase protein in the immune response signal
transduction.
KEYWORDS glycoprotein; immune response
FEATURE
1-17 #domain signal sequence #status predicted #label SIG\
18-467 #product Gram-negative bacteria-binding protein #status
predicted #label MAT\
119,182 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 467 #molecular-weight 52217 #checksum 7389
Query Match 82.5%; Score 47; DB 2; Length 467;
Best Local Similarity 62.5%; Pred. No. 5.12e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 299 RGNALYS 306
1111111
QY 175 RGESELY 182
RESULT 5
ENTRY B71378 #type complete
TITLE probable recF protein (recF) - syphilis spirochete
ORGANISM #formal_name Treponema pallidum subsp. pallidum #common_name
syphilis spirochete
DATE 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
17-Mar-1999
ACCESSIONS B71378
REFERENCE A71250
#authors Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.;
Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton,
R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod,
M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson,
D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald,
L.; Artiach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.;
Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey,
L.; Weidman, J.; Smith, H.O.; Venter, J.C.
#journal Science (1998) 281:375-388
#title Complete genome sequence of Treponema pallidum, the syphilis
spirochete.
#cross-references MUID:98332770
#accession B71378
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-403 #label COL
##cross-references GB:AE001185; GB:AE000520; NID:g3322254; PID:g3322258
##experimental_source strain Nichols
GENETICS
#gene TP0003
SUMMARY #length 403 #molecular-weight 47467 #checksum 3730
Query Match 80.7%; Score 46; DB 2; Length 403;
Best Local Similarity 75.0%; Pred. No. 8.59e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 101 RTESELY 108
1111111
QY 175 RGESELY 182
RESULT 6
ENTRY C70655 #type complete
TITLE probable monooxygenase - Mycobacterium tuberculosis (strain
H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS C70655

```

REFERENCE
#authors      A70500
               Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
               C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
               III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
               Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
               Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
               Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
               Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
               Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
               Selton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
               Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal      Nature (1998) 393:537-544
#title        Deciphering the biology of Mycobacterium tuberculosis from
               the complete genome sequence.
#cross-references MUID:98295987
#accession    C70455
#status       preliminary; nucleic acid sequence not shown;
               translation not shown
#molecule_type DNA
#residues     1-489 #label COL
#cross-references GB:283864; GB:AL123456; NID:g3261687; PID:e301250;
               PID:g1781102
#experimental_source strain H37Rv
GENETICS
#gene         Rv3854C
#summary      #length 489 #molecular-weight 55326 #checksum 7368
               Query Match 78.9%; Score 45; DB 2; Length 489;
               Best Local Similarity 50.0%; Pred. No. 1.43e+01;
               Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db          54 RSDSDMYT 61
            11111111
Qy          175 RGESELYT 182

RESULT      7
ENTRY
TITLE       H64073 #type complete
ORGANISM    DNA-directed RNA polymerase (EC 2.7.7.6) beta chain -
            Haemophilus influenzae (strain Rd KW20)
DATE        #formal_name Haemophilus influenzae
            18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
            30-Jun-1998
ACCESSIONS  H64073
REFERENCE    A64000
#authors     Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
            Kirkness, E.F.; Kervlavage, A.R.; Bult, C.J.; Tomb, J.F.;
            Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
            Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
            Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
            J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
            M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
            D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
            J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
            Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
            Science (1995) 269:456-512
            Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd.
            #cross-references MUID:95350630
            #status       preliminary; nucleic acid sequence not shown; translation not shown
            #accession    H64073
            #molecule_type DNA
            #residues     1-189 #label TIGR
            #cross-references GB:U32745; GB:L42023; NID:g1573617; PID:g1573624;
            TIGR:HI0628
            CLASSIFICATION #superfamily transcription initiation factor sigma E
            KEYWORDS       DNA binding; sigma factor; transcription initiation
            SUMMARY        #length 189 #molecular-weight 21791 #checksum 827
               Query Match 77.2%; Score 44; DB 2; Length 189;
               Best Local Similarity 75.0%; Pred. No. 2.36e+01;
               Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db          65 RGESELYT 72
            11111111
Qy          175 RGESELYT 182

RESULT      9
ENTRY
TITLE       S70210 #type complete
ORGANISM    rpoE protein homolog - Photobacterium sp.
DATE        #formal_name Photobacterium sp.
            12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change
            17-Mar-1999
ACCESSIONS  S70210
REFERENCE    S70210
#authors     Chi, E.; Bartlett, D.H.
            Mol. Microbiol. (1995) 17:713-726
            An rpoE-like locus controls outer membrane protein synthesis
            and growth at cold temperatures and high pressures in the
            deep-sea bacterium Photobacterium SS9.
            #cross-references MUID:9611491
            #accession    S70210
            #status       preliminary; nucleic acid sequence not shown;
            translation not shown
            #molecule_type DNA
            #residues     1-192 #label CHI
            #cross-references EMBL:L41667; NID:g77746; PID:g77747
            #note          the nucleotide sequence was submitted to the EMBL Data
            Library, April 1995
            CLASSIFICATION #superfamily transcription initiation factor sigma E
            SUMMARY        #length 192 #molecular-weight 21858 #checksum 4185
               Query Match 77.2%; Score 44; DB 2; Length 192;

```

```
Best Local Similarity 75.0%; Pred. No. 2.36e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 66 RGEAFYT 73
    ||||:||
Qy 175 RGESELYT 182

RESULT 10
ENTRY    S50834      #type complete
TITLE    emerin - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change
24-Sep-1998
ACCESSIONS S50834
REFERENCE   S50834
#authors   Bione, S.; Maestrini, E.; Rivella, S.; Mancini, M.; Regis,
S.; Romgo, G.; Toniolo, D.
#journal   Nature Genet. (1994) 8:323-327
#title     Identification of a novel X-linked gene responsible for
Emery-Dreifuss muscular dystrophy.
#accession S50834
#status    preliminary
#molecule_type mRNA
#residues  1-254 #label BIO
#cross-references EMBL:X82434; NID:g600618; PID:g600619
GENETICS
#gene      GDB:EMD
#map_position Xq28-Xq28
#cross-references GDB:119108; OMIM:310300
#map_position Xq28-Xq28
KEYWORDS   muscular dystrophy
SUMMARY    #length 254 #molecular-weight 28994 #checksum 630

Query Match 77.2%; Score 44; DB 2; Length 254;
Best Local Similarity 42.9%; Pred. No. 2.36e+01;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 68 RGDADMY 74
    ||:::|
Qy 175 RGESELY 181

RESULT 11
ENTRY    JQ1275      #type complete
TITLE    RAD57 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YB8119.10; protein YDR004w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
10-Jul-1998
ACCESSIONS JQ1275; S50985
REFERENCE   JQ1275
#authors   Kans, J.A.; Mortimer, R.K.
#journal   Gene (1991) 105:139-140
#title     Nucleotide sequence of the RAD57 gene of Saccharomyces
cerevisiae.
#cross-references MUID:92039012
#accession JQ1275
#molecule_type DNA
#residues  1-460 #label KAN
#cross-references GB:M5061; NID:g172344; PID:g172345
REFERENCE   S50976
#authors   Murphy, L.; Richards, C.; Gentles, S.; Harris, D.
#submission submitted to the EMBL Data Library, January 1995
#accession S50985
#molecule_type DNA
#residues  1-460 #label MUR
#cross-references EMBL:248008; NID:g642799; PID:g642809; MIPS:YDR004w
COMMENT    This protein participates in the repair of radiation-induced damage
to DNA in meiosis.
GENETICS
#gene      SGD:RAD57
#cross-references SGD:S0002411; MIPS:YDR004w
#map_position 4R
```

```
nucleus; P-loop
125-132      #region nucleotide-binding motif A (P-loop)\
221-226      #region nucleotide-binding motif B
SUMMARY      #length 460 #molecular-weight 52247 #checksum 7074

Query Match 77.2%; Score 44; DB 2; Length 460;
Best Local Similarity 71.4%; Pred. No. 2.36e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 414 RGEAHLV 420
    ||||:|
Qy 175 RGESELY 181

RESULT 12
ENTRY    B69941      #type complete
TITLE    ATP-dependent helicase homolog yprA - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE     05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
ACCESSIONS B69941
REFERENCE   A69580
#authors   Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Jones, B.; Karamata, D.;
Kasahara, Y.; Klaert-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roché, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weltzenegger, I.;
Winters, K.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
#journal   Nature (1997) 390:249-256
#title     The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession B69941
#status    preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues  1-749 #label KUN
#cross-references GB:299115; GB:AL009126; NID:g2634478; PID:el183668;
PID:g2634641
#experimental_source strain 168

GENETICS
#gene      yprA
CLASSIFICATION #superfamily unassigned DEAD/H box helicases; DEAD/H box
```



```
KEYWORDS      helicase homology
FEATURE
76-387        ATP; P-Loop
76-83         #domain DEAD/H box helicase homology #label DEAD\
              #region nucleotide-binding motif A (P-loop) #status
              atypical\
181-186       #region nucleotide-binding motif B\
185-188       #region DEXH motif
SUMMARY       #length 749 #molecular-weight 84589 #checksum 3628

Query Match   77.2%; Score 44; DB 2; Length 749;
Best Local Similarity 75.0%; Pred. No. 2.36e+01;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 50 RGIDELYT 57
   || ||||
Qy 175 RGESELYT 182

RESULT 13
ENTRY   A53034 #type complete
TITLE   gag polyprotein - human immunodeficiency virus type 1 (strain
         Ant70)
ORGANISM #formal_name human immunodeficiency virus type 1, HIV-1
DATE     03-May-1994 #sequence_revision 03-May-1994 #text_change
         07-Feb-1997
ACCESSIONS A53034
REFERENCE  A53034
AUTHORS    Vanden Haesevelde, M.; Decourt, J.L.; De Leys, R.J.;
           Vanderborght, B.; van der Groen, G.; van Heuverswijn, H.;
           Saman, E.
JOURNAL    J. Virol. (1994) 68:1586-1596
TITLE      Genomic cloning and complete sequence analysis of a highly
           divergent African human immunodeficiency virus isolate.
ACCESSION  A53034
STATUS     preliminary
RESIDUES   1-863 #label VAN
CROSS-REFERENCES GB:L02587
KEYWORDS   polypeptide
SUMMARY    #length 863 #molecular-weight 97102 #checksum 7006

Query Match   77.2%; Score 44; DB 2; Length 863;
Best Local Similarity 62.5%; Pred. No. 2.36e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 426 RQSGSLYA 433
   || |||
Qy 175 RGESELYT 182

RESULT 14
ENTRY   T00347 #type fragment
TITLE   hypothetical protein KIAA0592 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE     01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
         01-Feb-1999
ACCESSIONS T00347
REFERENCE  Z14086
AUTHORS    Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani,
           H.; Nomura, N.; Ohara, O.
JOURNAL    DNA Res. (1998) 5:31-39
TITLE      Prediction of the coding sequences of unidentified human
           genes. IX. The complete sequences of 100 new cDNA clones
           from brain which can code for large proteins in vitro.
ACCESSION  T00347
STATUS     preliminary; translated from GB/EMBL/DBJ
MOLECULE_TYPE mRNA
RESIDUES   1-1353 #label NAG
CROSS-REFERENCES EMBL:AB011164; NID:d1185434; PID:d1026448
EXPERIMENTAL_SOURCE brain; clone HJ2807
GENETICS
```

```
#note      KIRA0592
SUMMARY    #length 1353 #checksum 6126

Query Match   77.2%; Score 44; DB 2; Length 1353;
Best Local Similarity 57.1%; Pred. No. 2.36e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1136 RGEADLF 1142
   |||||
Qy 175 RGESELYT 181

RESULT 15
ENTRY   JQ1896 #type complete
TITLE   hypothetical 191.1k protein - raspberry bushy dwarf virus
ORGANISM #formal_name raspberry bushy dwarf virus
DATE     30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
         30-Sep-1993
ACCESSIONS JQ1896
REFERENCE  JQ1896
AUTHORS    Ziegler, A.; Natsuaki, T.; Mayo, M.A.; Jolly, C.A.; Murant,
           A.F.
JOURNAL    J. Gen. Virol. (1992) 73:3213-3218
TITLE      The nucleotide sequence of RNA-1 of raspberry bushy dwarf
           virus.
ACCESSION  JQ1896
MOLECULE_TYPE mRNA
RESIDUES   1-1694 #label ZIE
GENETICS   #map_position segment RNA-1
KEYWORDS   RNA replication
SUMMARY    #length 1694 #molecular-weight 191098 #checksum 1747

Query Match   77.2%; Score 44; DB 2; Length 1694;
Best Local Similarity 50.0%; Pred. No. 2.36e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1084 RTDNDLYT 1091
   | :|||
Qy 175 RGESELYT 182

Search completed: Thu Jul 8 18:25:59 1999
Job time : 10 secs.
```

THIS PAGE BLANK (USPTO)

WPSREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:24:49 1999; MasPar time 3.18 Seconds
Tabular output not generated. 71.035 Million cell updates/sec

Title: >US-09-041-236-2
Description: (175-182) from US09041236.pep (10 of 45)
Perfect Score: 57
Sequence: 1 RGESELYT 8

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 22.747; Variance 21.933; scale 1.037

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	47	82.5	31	1	PLD_BRAOL PHOSPHOLIPASE D (EC 3.1.1.4)	1.82e+00
2	46	80.7	403	1	REC_TREPA REC_F PROTEIN.	3.21e+00
3	45	78.9	1343	1	RPOB_HAEIN DNA-DIRECTED RNA POLYM	5.60e+00
4	44	77.2	189	1	RPOE_HAEIN RNA POLYMERASE SIGMA-E	9.69e+00
5	44	77.2	254	1	EMD_HUMAN EMERIN	9.69e+00
6	44	77.2	397	1	GLGC_STRCO GLUCOSE-1-PHOSPHATE AD	9.69e+00
7	44	77.2	460	1	RAS7_YEAST DNA REPAIR PROTEIN RAD	9.69e+00
8	44	77.2	749	1	YPRA_BACSU HYPOTHETICAL HELICASE	9.69e+00
9	43	75.4	78	1	RL38_YEAST 60S RIBOSOMAL PROTEIN	1.66e+01
10	43	75.4	104	1	Y079_NPVOP HYPOTHETICAL 12.2 KD P	1.66e+01
11	43	75.4	191	1	RPOE_ECOLI RNA POLYMERASE SIGMA-E	1.66e+01
12	43	75.4	191	1	RPOE_SALTY RNA POLYMERASE SIGMA-E	1.66e+01
13	43	75.4	193	1	RPSH_PSEAE RNA POLYMERASE SIGMA-H	1.66e+01
14	43	75.4	211	1	SSPA_ECOLI STRINGENT STARVATION P	1.66e+01
15	43	75.4	291	1	AMPN_ARCFU METHIONINE AMINOPEPTID	1.66e+01
16	43	75.4	390	1	DAFL_MOUSE COMPLEMENT DECAV-ACCEL	1.66e+01
17	43	75.4	415	1	SACE_ERWAM LEVANSUCRASE (EC 2.4.1.1)	1.66e+01
18	43	75.4	442	1	VANI_CANAL VANADATE RESISTANCE PR	1.66e+01
19	43	75.4	660	1	CCMF_BRAJA CYTOCHROME C-TYPE BIOG	1.66e+01
20	43	75.4	682	1	KPCG_BOVIN PROTEIN KINASE C, GAMM	1.66e+01
21	43	75.4	697	1	KPCG_MOUSE PROTEIN KINASE C, GAMM	1.66e+01
22	43	75.4	697	1	KPCG_RABIT PROTEIN KINASE C, GAMM	1.66e+01
23	43	75.4	697	1	KPCG_HUMAN PROTEIN KINASE C, GAMM	1.66e+01

24	43	75.4	711	1	PTIA_ECOLI PHOSPHOENOLPYRUVATE-PR	1.66e+01
25	42	73.7	350	1	ENP2_TORCA ELECTROMOTOR NEURON-AS	2.81e+01
26	42	73.7	363	1	NIR_PSEAR COPPER-CONTAINING NITR	2.81e+01
27	42	73.7	378	1	APPB_ECOLI CYTOCHROME BD-II OXIDA	2.81e+01
28	42	73.7	451	1	PHOQ_SALTY VIRULANCE SENSOR PROTE	2.81e+01
29	42	73.7	486	1	PHOQ_ECOLI SENSOR PROTEIN PHOQ (E	2.81e+01
30	42	73.7	506	1	Y619_METJA HYPOTHETICAL PROTEIN M	2.81e+01
31	42	73.7	536	1	SR54_YARLI SIGNAL RECOGNITION PAR	2.81e+01
32	42	73.7	537	1	YHES_ECOLI HYPOTHETICAL ABC TRANS	2.81e+01
33	42	73.7	700	1	CAOQ_RAT PRISTANOL-COA OXIDASE	2.81e+01
34	42	73.7	700	1	KPC2_DROME PROTEIN KINASE C, EYE	2.81e+01
35	42	73.7	721	1	ENPL_TORCA ELECTROMOTOR NEURON-AS	2.81e+01
36	42	73.7	740	1	PSAB_ANAVA PHOTOSYSTEM I P700 CHL	2.81e+01
37	42	73.7	1043	1	PLID_MOUSE PHOSPHATIDYLINOSITOL 3	2.81e+01
38	42	73.7	1044	1	PLID_HUMAN PHOSPHATIDYLINOSITOL 3	2.81e+01
39	42	73.7	1284	1	ATI_COMPX A-TYPE INCLUSION PROTE	2.81e+01
40	42	73.7	1403	1	PRO_DROME PROTEIN PROSPERO	2.81e+01
41	42	73.7	2464	1	MAPB_MOUSE MICROTUBULE-ASSOCIATED	2.81e+01
42	42	73.7	2468	1	MAPB_HUMAN MICROTUBULE-ASSOCIATED	2.81e+01
43	42	73.7	2774	1	MAPA_RAT MICROTUBULE-ASSOCIATED	2.81e+01
44	42	73.7	2805	1	MAPA_HUMAN MICROTUBULE-ASSOCIATED	2.81e+01
45	42	73.7	3051	1	YNX3_CAEL HYPOTHETICAL PROTEIN T	2.81e+01

ALIGNMENTS

RESULT 1	PLD_BRAOL	STANDARD:	PRT:	31 AA.
ID	AC P55939;			
DT	01-NOV-1997 (REL. 35; CREATED)			
DT	01-NOV-1997 (REL. 35; LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35; LAST ANNOTATION UPDATE)			
DE	PHOSPHOLIPASE D (EC 3.1.4.4) (PLD) (CHOLINE PHOSPHATASE)			
DE	(PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D) (FRAGMENT).			
OS	BRASSICA OLERACEA (CAULIFLOWER).			
OC	EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;			
OC	EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;			
OC	CAPPARALES; BRASSICACEAE; BRASSICA.			
RN	[1]			
RP	SEQUENCE.			
RC	STRAIN=CV. CAPITATA; TISSUE=LEAF;			
RX	MEDLINE: G3357261.			
RA	ABOUSALHAM A., RIVIERE M., TEISSERE M., VERGER R.;			
RT	"Improved purification and biochemical characterization of			
RT	phospholipase D from cabbage."			
RL	BIOCHIM. BIOPHYS. ACTA 1158:1-7(1993).			
CC	-!- FUNCTION: PLAYS AN IMPORTANT ROLE IN VARIOUS CELLULAR PROCESSES.			
CC	OPTIMAL PH DEPENDS ON CALCIUM CONCENTRATION			
CC	-!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A			
CC	PHOSPHATIDATE.			
CC	-!- COFACTOR: CALCIUM.			
CC	-!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.			
KW	HYDROLASE; LIPID DEGRADATION; CALCIUM.			
FT	NON_TER 31			
SQ	SEQUENCE 31 AA; 3388 MW; EIBBF6FB CRC32;			
Query Match	82.5%;	Score 47;	DB 1;	Length 31;
Best Local Similarity	62.5%;	Pred. No. 1.82e+00;		
Matches	5;	Conservative	3;	Mismatches 0; Indels 0; Gaps 0;
Db	9 KGETELYA 16			
QY	175 RGESELYT 182			
	:			
RESULT 2	REC_TREPA	STANDARD:	PRT:	403 AA.
ID	AC O83049;			
DT	15-DEC-1998 (REL. 37; CREATED)			
DT	15-DEC-1998 (REL. 37; LAST SEQUENCE UPDATE)			
DE	REC_F PROTEIN.			
DE	REC_F PROTEIN.			
GN	REC_F OR TP0003.			

OS TREPONEMA PALLIDUM.
 OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; TREPONEMA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NICHOLS;
 RX MEDLINE; 98332770.
 RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
 RA DODSON R., GINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
 RA DOSBERGEN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTERBACK T.,
 RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
 RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
 RA VENTER J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 spirochete.";
 RL SCIENCE 281:375-388(1998).
 CC -!- FUNCTION: THE RCF PROTEIN IS INVOLVED IN DNA METABOLISM; IT IS
 CC REQUIRED FOR DNA REPLICATION AND NORMAL SOS INDUCIBILITY. RCF
 CC BINDS PREFERENTIALLY TO SINGLE-STRANDED, LINEAR DNA. IT ALSO SEEMS
 CC TO BIND ATP (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE RCF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE001185; G3322258; -.
 DR TIGR; TP0003; -.
 DR PROSITE; PS00617; RCF_1; 1.
 DR PROSITE; PS00618; RCF_2; FALSE_NEG.
 KW DNA DAMAGE; DNA REPLICATION; DNA-BINDING; SOS RESPONSE; DNA REPAIR;
 KW ATP-BINDING.
 FT NP_BIND 76 83 ATP (POTENTIAL).
 SQ SEQUENCE 403 AA; 47467 MW; 373EBF18 CRC32;

Query Match 80.7%; Score 46; DB 1; Length 403;
 Best Local Similarity 75.0%; Pred. No. 3.21e+00;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 101 RTESELYA 108
 | | | | | | | |
 Qy 175 RGESELYT 182

RESULT 3
 ID RPOB_HAEIN STANDARD; PRT; 1343 AA.
 AC P43738;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DE DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6) (TRANSCRIPTASE
 DE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT).
 GN RPOB OR HI0515.
 OS HAEMOPHILUS INFLUENZAE.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
 OC HAEMOPHILUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20;
 RX MEDLINE; 95350630.
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
 RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
 RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
 RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 influenzae Rd.";
 RL SCIENCE 269:496-512(1995).
 CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
 CC THEN IS RELEASED. THIS SIGMA FACTOR IS INVOLVED IN HEAT SHOCK AND
 CC OXIDATIVE STRESS RESPONSE: IT IS BELIEVED TO CONTROL PROTEIN
 CC PROCESSING IN THE CYTOPLASMIC COMPARTMENT (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY. ECF SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 influenzae Rd.";
 RL SCIENCE 269:496-512(1995).
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE --> N PYROPHOSPHATE +
 CC RNA(N).
 CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
 CC BETA' CHAIN.
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U32733; G1573496; -.
 DR TIGR; HI0515; -.
 DR PROSITE; PS01166; RNA_POL_BETA; 1.
 DR PRAM; PF00562; RNA_POL_B; 1.
 KW TRANSFERASE; TRANSCRIPTION; DNA-DIRECTED RNA POLYMERASE.
 SQ SEQUENCE 1343 AA; 149783 MW; A6F024D1 CRC32;

Query Match 78.9%; Score 45; DB 1; Length 1343;
 Best Local Similarity 85.7%; Pred. No. 5.60e+00;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 640 RGESELY 646
 | | | | | | | |
 Qy 175 RGESELY 181

RESULT 4
 ID RPOE_HAEIN STANDARD; PRT; 189 AA.
 AC P47950;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE RNA POLYMERASE SIGMA-E FACTOR (SIGMA-24).
 GN RPOE OR HI0628
 OS HAEMOPHILUS INFLUENZAE.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
 OC HAEMOPHILUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20;
 RX MEDLINE; 95350630.
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
 RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
 RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
 RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 influenzae Rd.";
 RL SCIENCE 269:496-512(1995).
 CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
 CC THEN IS RELEASED. THIS SIGMA FACTOR IS INVOLVED IN HEAT SHOCK AND
 CC OXIDATIVE STRESS RESPONSE: IT IS BELIEVED TO CONTROL PROTEIN
 CC PROCESSING IN THE CYTOPLASMIC COMPARTMENT (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY. ECF SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U32745; G1573624; -;
DR TIGR; H10628; -;
DR PROSITE; PS01063; SIGMA70_ECF; 1.
DR PFAM; PF00776; Sigma70_ECF; 1.
KW TRANSCRIPTION REGULATION; SIGMA FACTOR; DNA-DIRECTED RNA POLYMERASE;
DNA-BINDING; HEAT SHOCK.
FT DOMAIN 48 61 POLYMERASE CORE BINDING (POTENTIAL).
FT DNA_BIND 156 175 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 189 AA; 21791 MW; 580453B3 CRC32;

Query Match 77.2%; Score 44; DB 1; Length 189;
Best Local Similarity 75.0%; Pred. No. 9.69e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 65 RGSFAFYT 72
|||:|
Qy 175 RGESELYT 182

RESULT 5
ID EMD_HUMAN STANDARD; PRT; 254 AA.
AC P50402;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE EMERIN.
GN EMD OR STA OR EMD.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]

RP SEQUENCE FROM N.A.
RC TISSUE-TERATOCARCINOMA;
RX MEDLINE; 95201795.
RA BIONE S., MAESTRINI E., RIVELLA S., MANCINI M., REGIS S.,
RA ROMEO G., TONIOLO D.;
RT "Identification of a novel X-linked gene responsible for
RT Emery-Dreifuss muscular dystrophy."
RL NAT. GENET. 8:323-327(1994).
[2]
RP SEQUENCE FROM N.A.
RA CHEN E.Y., ZOLLO M., MAZZARELLA R.A., CICCOCICOLA A., CHEN C.N.,
RA ZUO L., HEINER C., BURROGH F.W., RIPETTO M., SCHLESSINGER D.,
RA D'URSO M.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96209922.
RA YANADA T., KOBAYASHI T.;
RT "A novel emerin mutation in a Japanese patient with Emery-Dreifuss
RT muscular dystrophy."
RL HUM. GENET. 97:693-694(1996).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96121582.
RA BIONE S., SMALL K., AKSMANOVIC M.A., D'URSO M., CICCOCICOLA A.,
RA MERLINI L., MORANDI L., KRESS W., YATES J.R.W., WARREN S.T.,
RA TONIOLO D.;
RT "Identification of new mutations in the Emery-Dreifuss muscular
RT dystrophy gene and evidence for genetic heterogeneity of the
RT disease."
RL HUM. MOL. GENET. 4:1859-1863(1995).

CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE, HEART, COLON, TESTIS, OVARY
CC AND PANCREAS.
CC -!- DISEASE: DEFECTS IN EMD ARE A CAUSE OF EMERY-DREIFUSS MUSCULAR
CC DYSTROPHY (EMDM), AN X-LINKED DISORDER CHARACTERIZED BY EARLY

CC CONTRACTURES, MUSCLE WASTING AND WEAKNESS AND CARDIOMYOPATHY.
CC -!- DATABASE: NAME=EMD db; NOTE="EMD mutation database";
CC WWW="http://www.path.cam.ac.uk/emd/";
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X82434; G600619; -;
DR EMBL; L44140; G1203970; -;
DR EMBL; D64111; G1794156; -;
DR EMBL; X86810; E1154169; -;
DR MIN; 310300; -;
KW TRANSMEMBRANE; DISEASE MUTATION.
FT DOMAIN 192 199 POLY-SER.
FT TRANSVMEM 223 243 POTENTIAL.
FT VARIANT 54 54 S -> F (IN EDM).
FT VARIANT 183 183 P -> H (IN EDM).
FT VARIANT 183 183 P -> T (IN EDM).
SQ SEQUENCE 254 AA; 28993 MW; DB100CA CRC32;

Query Match 77.2%; Score 44; DB 1; Length 254;
Best Local Similarity 42.9%; Pred. No. 9.69e+00;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 68 RGDADMY 74
||:|:|
Qy 175 RGESELY 181

RESULT 6
ID GLGC_STRCO STANDARD; PRT; 397 AA.
AC P72394;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GLUCOSE-1-PHOSPHATE ADENYLYLTRANSFERASE (EC 2.7.7.27) (ADP-GLUCOSE
DE SYNTHASE) (ADP-GLUCOSE PYROPHOSPHYLASE).
GN GLGC.
OS STREPTOMYCES COELICOLOR.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 98062210.
RA MARTIN M.C., SCHNEIDER D., BRUTON C.J., CHATER K.F., HARDISSON C.;
RT "A glcC gene essential only for the first of two spatially distinct
RT phases of glycogen synthesis in Streptomyces coelicolor A3(2).";
RL J. BACTERIOL 179:7784-7789(1997).
CC -!- CATALYTIC ACTIVITY: ADP + ALPHA-D-GLUCOSE 1-PHOSPHATE =
CC PYROPHOSPHATE + ADP-GLUCOSE.
CC -!- PATHWAY: FIRST STEP IN GLYCOGEN BIOSYNTHESIS.
CC -!- SIMILARITY: TO OTHER BACTERIAL AND PLANTS GLUCOSE-1-PHOSPHATE
CC ADENYLYLTRANSFERASES.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X89733; E267370; -;
DR PROSITE; PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE; PS00809; ADP_GLC_PYROPHOSPH_2; 1.
DR PROSITE; PS00810; ADP_GLC_PYROPHOSPH_3; 1.
DR PFAM; PF00483; NTP_transferase; 1.

KW GYCOGEN BIOSYNTHESIS; TRANSFERASE; NUCLEOTIDYLTRANSFERASE.
SQ SEQUENCE 397 AA; 42579 MW; 618C15A8 CRC32;

Query Match 77.2%; Score 44; DB 1; Length 397;
Best Local Similarity 71.4%; Pred. No. 9.69e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 230 RGEAALY 236
|||: ||
QY 175 RGESELY 181

RESULT 7
ID RA57-YEAST STANDARD; PRT; 460 AA.
AC P25301;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DNA REPAIR PROTEIN RAD57.
GN RAD57 OR YD8004 OR YD8119.10.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
[1]
RN SEQUENCE FROM N.A.
RP EMBL; 72039012.
RX MEDLINE; 92039012.
RA KANS J.A., MORTIMER R.K.;
RT "Nucleotide sequence of the RAD57 gene of *Saccharomyces cerevisiae*.";
RL GENE 105:139-140(1991).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN-S288C / AB972;
RA MURPHY L., RICHARDS C., GENTLES S., HARRIS D., BARRELL B.G.,
RA RAJANDREAM M.A.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: PARTICIPATES IN THE REPAIR OF X-RAY-INDUCED DAMAGE TO
CC DNA AND IN MEIOSIS. IT MAY ACT IN PART BY STABILIZING A REPAIR
CC COMPLEX OF OTHER RAD GENES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M55061; G172345; -.
DR EMBL; 248008; G542809; -.
DR PIR; JQ1275; JQ1275.
DR SGD; L0001577; RAD57.
KW DNA DAMAGE; DNA REPAIR; ATP-BINDING; MEIOSIS; NUCLEAR PROTEIN.
FT NP_BIND 125 132
FT SITE 125 132 ATP (POTENTIAL).
SQ SEQUENCE 460 AA; 52247 MW; 2B9FD07F CRC32;

Query Match 77.2%; Score 44; DB 1; Length 460;
Best Local Similarity 71.4%; Pred. No. 9.69e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 414 RGEAHLV 420
|||: ||
QY 175 RGESELY 181

RESULT 8
ID YPRA_BACSU STANDARD; PRT; 749 AA.
AC P50830;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL HELICASE IN PONA-COTD INTERGENIC REGION.
GN YPRA.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE; 96349105.
RA SOROKIN A.V., AZEVEDO V., ZUMSTEIN E., GALLERON N., EHRLICH S.D.,
RA SERROR P.;
RT "Sequence analysis of the *Bacillus subtilis* chromosome region between
RT the *serA* and *kdg* loci cloned in a yeast artificial chromosome.";
RL MICROBIOLOGY 142:2005-2016(1996).
CC -1- SIMILARITY: WITH OTHER ATP DEPENDENT HELICASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L47838; G1146178; -.
CC EMBL; 299115; E1183668; -.
DR SUBTILIST; BG11451; YPRA.
DR PFAM; PF00271; helicase.C.1.
KW HYPOTHETICAL PROTEIN; HELICASE; ATP-BINDING.
FT NP_BIND 76 83
FT SITE 185 188 ATP (BY SIMILARITY).
FT DEVH BOX.
SQ SEQUENCE 749 AA; 84589 MW; 0C9E0C9A CRC32;

Query Match 77.2%; Score 44; DB 1; Length 749;
Best Local Similarity 75.0%; Pred. No. 9.69e+00;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 50 RGIDELYT 57
|||: |||
QY 175 RGESELYT 182

RESULT 9
ID RL38-YEAST STANDARD; PRT; 78 AA.
AC P49167;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 60S RIBOSOMAL PROTEIN L38.
GN RPL38 OR YLR325C OR L8543.2.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J.,
RA KUCABA T., HALLSWORTH K., HANKINS J., HILLIER L., JIER M.,
RA JOHNSON D., JOHNSON L., LANGSTON Y., LATREILLE P., LE T.,
RA MARDIS E., MENEZES S., MILLER N., NHAN M., PAULEY A., PELUSO D.,
RA RIFKIN L., RILES L., TAICH A., TREVASKIS E., VIGNATI D.,
RA WILCOX L., WOHLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: BELONGS TO THE L38E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```
CC -----
DR EMBL; U20618; G662127; -
KW RIBOSOMAL PROTEIN. RPL38.
SQ SEQUENCE 78 AA; 8826 MW; 6243348F CRC32;

Query Match      75.4%; Score 43; DB 1; Length 78;
Best Local Similarity 75.0%; Pred. No. 1.66e+01;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 46 RGSSSLYT 53
   |||
Oy 175 RGESELYT 182

RESULT 10
ID Y079_NPVOP STANDARD; PRT; 104 AA.
AC O1032;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 12.2 KD PROTEIN (ORF82).
OS ORGYIA PSEUDOTSUGATA MULTICAPSID POLYHEDROSIS VIRUS (OPMPNV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
CC NUCLEOPOLYHEDROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97217300.
RA AURENS C.H., RUSSELL R.R., FUNK C.J., EVANS J., HARWOOD S.,
RA ROHRMANN G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";
RL VIROLOGY 229:381-399(1997).
CC -!- SIMILARITY: TO CORRESPONDING ORF IN ACMPV.
CC -!- SIMILARITY: STRONG, TO E.COLI YHBQ.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75930; G1911328; -
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 104 AA; 12205 MW; BBC0A8D2 CRC32;

Query Match      75.4%; Score 43; DB 1; Length 104;
Best Local Similarity 50.0%; Pred. No. 1.66e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 16 RDDGQLYT 23
   |::|||
Oy 175 RGESELYT 182

RESULT 11
ID RPOE_ECOLI STANDARD; PRT; 191 AA.
AC P34086;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE SIGMA-E FACTOR (SIGMA-24).
GN RPOE OR SIGE.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RA NASHIMOTO H.;
RT "Non-ribosomal proteins affecting the assembly of ribosomes in
```

```
RT Escherichia coli.";
RL (IN) NIERHAUS K.H. (EDS.);
RL THE TRANSLATIONAL APPARATUS, PP.185-195, PLENUM PRESS,
RL NEW YORK (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA MISSIAKAS D., RAINA S.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA NASHIMOTO H., SAITO N.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL SCIENCE 277:1453-1474(1997).
RN [5]
RP SEQUENCE OF 171-191 FROM N.A.
RC STRAIN=K12, VD1870;
RX MEDLINE; 95286510.
RA YU H., SCHURR M.J., DERETIC V.;
RT "Functional equivalence of Escherichia coli sigma E and Pseudomonas
RT aeruginosa Algu: E. coli rpoE restores mucooidy and reduces
RT sensitivity to reactive oxygen intermediates in algu mutants of P.
RT aeruginosa.";
RL J. BACTERIOL. 177:3259-3268(1995).
RN [6]
RP IDENTIFICATION, AND SEQUENCE OF 40-56.
RX MEDLINE; 95196747.
RA ROUVIERE P.E., DE LAS PENAS A., MECSAS J., LU C.Z., RUDD K.E.,
RA GROSS C.A.;
RT "rpoE, the gene encoding the second heat-shock sigma factor, sigma E,
RT in Escherichia coli.";
RL EMBO J. 14:1032-1042(1995).
RN [7]
RP IDENTIFICATION.
RX MEDLINE; 94329538.
RA LONETTO M.A., BROWN K.L., RUDD K.E., BUTTNER M.J.;
RT "Analysis of the Streptomyces coelicolor sigE gene reveals the
RT existence of a subfamily of eubacterial RNA polymerase sigma factors
RT involved in the regulation of extracytoplasmic functions.";
RL PROC. NATL. ACAD. SCI. U.S.A. 91:7573-7577(1994).
RN [8]
RP IDENTIFICATION.
RX MEDLINE; 95050230.
RA MARTIN D.W., SCHURR M.J., YU H., DERETIC V.;
RT "Analysis of promoters controlled by the putative sigma factor Algu
RT regulating conversion to mucooidy in Pseudomonas aeruginosa:
RT relationship to sigma E and stress response.";
RL J. BACTERIOL. 176:6688-6696(1994).
RN [9]
RP IDENTIFICATION, AND LOSS OF ACTIVITY MUTANTS.
RX MEDLINE; 95196748.
RA RAINA S., MISSIAKAS D., GEORGOPOULOS C.P.;
RT "The rpoE gene encoding the sigma E (sigma 24) heat shock sigma
RT factor of Escherichia coli.";
RL EMBO J. 14:1043-1055(1995).
CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED. THIS SIGMA FACTOR IS INVOLVED IN HEAT SHOCK AND
CC OXIDATIVE STRESS RESPONSE; IT IS BELIEVED TO CONTROL PROTEIN
CC PROCESSING IN THE EXTRACYTOPLASMIC COMPARTMENT.
CC -!- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY. ECF SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```


KW TRANSCRIPTION REGULATION; SIGMA FACTOR; DNA-DIRECTED RNA POLYMERASE;
KW DNA-BINDING. 49 62 POLYMERASE CORE BINDING (POTENTIAL).
FT DOMAIN 157 176 H-T-H MOTIF (BY SIMILARITY).
FT DNA_BIND 157 176
SQ SEQUENCE 193 AA; 22196 MW; 4A1640DA CRC32;

Query Match 75.4%; Score 43; DB 1; Length 193;
Best Local Similarity 62.5%; Pred. No. 1.66e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 66 RGSAPFYT 73
QY 175 RGESELYT 182

RESULT 14

ID SSPA_ECOLI STANDARD; PRT; 211 AA.
AC P05938;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE STRINGENT STARVATION PROTEIN A.
GN SSPA OR SSP OR POG.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.

[1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 87146422.
RA SERIZAWA H., FUKUDA R.;
RT "Structure of the gene for the stringent starvation protein of
RT Escherichia coli";
RL NUCLEIC ACIDS RES. 15:1153-1163(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., ROSEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL SCIENCE 277:1453-1474(1997).
RN [3]
RP SEQUENCE OF 1-12.
RC STRAIN-K12 / EMG2;
RX MEDLINE; 97443975.
RA LINK A.J., ROBISON K., CHURCH G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12";
RL ELECTROPHORESIS 18:1259-1313(1997).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE; 94293773.
RA WILLIAMS M.D., OUYANG T.X., FLICKINGER M.C.;
RT "Starvation-induced expression of SspA and SspB: the effects of a
RT null mutation in sspA on Escherichia coli protein synthesis and
RT survival during growth and prolonged starvation";
RL MOL. MICROBIOL. 11:1029-1043(1994).

CC -!- FUNCTION: FORMS AN EQUIMOLAR COMPLEX WITH THE RNA POLYMERASE
CC HOLOENZYME (RNAP) BUT NOT WITH THE CORE ENZYME, IT IS SYNTHESIZED
CC PREDOMINANTLY WHEN CELLS ARE EXPOSED TO AMINO ACID STARVATION, AT
CC WHICH TIME IT ACCOUNTS FOR OVER 50% OF THE TOTAL PROTEIN
CC SYNTHESIZED. IT IS INVOLVED IN THE TRANSITION FROM P1 EARLY TO P1
CC LATE GENE EXPRESSION. RNK AND SSPA CAN FUNCTIONALLY REPLACE
CC P.AERUGINOSA ALGINATE REGULATORY GENE ALGR2.
CC -!- INDUCTION: BY AMINO ACID STARVATION
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. HSP26 FAMILY. STRONG,
CC TO OTHER BACTERIAL SSPA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X05088; G42998;
DR EMBL; U18997; G606168;
DR EMBL; AE000402; G1789624;
DR PIR; A26422; RGCSS.
DR SWISS-2DPAGE; P05838; COLI.
DR ECO2DBASE; D027.1; 6TH EDITION.
DR ECOGENE; EG10977; SSPA.
DR PAM; PF00043; gluts; 1.
FT INIT_MET 0
SQ SEQUENCE 211 AA; 24173 MW; 5457EC4D CRC32;

Query Match 75.4%; Score 43; DB 1; Length 211;
Best Local Similarity 85.7%; Pred. No. 1.66e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 95 RGSRLY 101
QY 175 RGESELY 181

RESULT 15

ID AMPM_ARCFU STANDARD; PRT; 291 AA.
AC O28438;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE METHIONINE AMINOPEPTIDASE (EC 3.4.11.18) (MAP) (PEPTIDASE M).
GN MAP OR AF1840.
OS ARCHAEoglobus fulgidus.
OC ARCHAEA; EURYARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae;
OC ARCHAEoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V-16 / DSM 4304 / ATCC 49558;
RX MEDLINE; 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,
RA FLEISCHMANN R.D., QUACRENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBEK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus";
RL NATURE 390:364-370(1997).

CC -!- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
CC PROTEINS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE
CC + PEPTIDE.
CC -!- COFACTOR: COBALT (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C; ALSO KNOWN AS THE
CC MAP FAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE000976; G2648708;
DR TIGR; AF1840;
DR PROSITE; PS01202; MAP_2; 1.

```

DR  PFM; PF00557; pep_M24; 1.
DR  HSP; P56218; LXGO.
KW  HYDROLASE; AMINOPEPTIDASE; COBALT.
FT  METAL      85  85  COBALT (BY SIMILARITY).
FT  METAL      96  96  COBALT (BY SIMILARITY).
FT  METAL     155  155  COBALT (BY SIMILARITY).
FT  METAL     188  188  COBALT (BY SIMILARITY).
FT  METAL     277  277  COBALT (BY SIMILARITY).
SQ  SEQUENCE  291 AA; 31691 MW; 474065f8 CRC32;

Query Match      75.4%; Score 43; DB 1; Length 291;
Best Local Similarity 62.5%; Pred. No. 1.66e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db  201 RGECEIYS 208
QY  175 RGESELYT 182

Search completed: Thu Jul  8 18:24:57 1999
Job time : 8 secs.

```

W P E R E H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:25:16 1999; MasPar time 7.91 Seconds

Tabular output not generated. 55.201 Million cell updates/sec

Title: >US-09-041-236-2
Description: (175-182) from US09041236.pep (10 of 45)
Perfect Score: 57

Sequence: 1 RGESELYT 8

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 21.983; Variance 22.750; scale 0.966

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Match	Length	ID	Description	Pred. No.
1	57	100.0	393	11	088371 SEMAPHORIN L (FRAGMENT)	1.15e-02
2	57	100.0	666	4	075326 SEMAPHORIN L	1.15e-02
3	50	87.7	250	2	051348 CHAPERONIN, PUTATIVE.	7.60e-01
4	49	86.0	67	2	056923 HYPOTHETICAL PROTEIN (1.35e+00
5	49	86.0	131	2	056851 PRE-LIPOPROTEIN VIRG P	1.35e+00
6	49	86.0	131	2	068694 PRE-LIPOPROTEIN VIRG H	1.35e+00
7	47	82.5	467	5	017233 P50 PROTEIN PRECURSOR.	4.11e+00
8	45	78.9	489	2	096223 HYPOTHETICAL 55.3 KD P	1.22e+01
9	45	78.9	771	5	046307 EG:8D8.4 PROTEIN.	1.22e+01
10	44	77.2	192	2	057093 RPOE, LEPA AND ORF 2,	2.06e+01
11	44	77.2	356	13	093336 HEPARAN SULFATE 2-SULF	2.06e+01
12	44	77.2	356	4	075036 KIAA0448 PROTEIN.	2.06e+01
13	44	77.2	810	10	082549 PHOSPHOLIPASE D1 (EC 3	2.06e+01
14	44	77.2	812	10	049981 PHOSPHOLIPASE D (EC 3.	2.06e+01
15	44	77.2	812	10	082548 PHOSPHOLIPASE D2 (EC 3	2.06e+01
16	44	77.2	863	14	077377 ENVELOPE PROTEIN GP120	2.06e+01
17	44	77.2	1694	14	005983 RNA1 POLYPROTEIN.	2.06e+01
18	43	75.4	347	10	049623 3'(2'),5'-BISPHOSPHATE	3.47e+01
19	43	75.4	356	11	088464 HEPARAN SULFATE 2-SULF	3.47e+01
20	43	75.4	356	11	008889 HEPARAN SULFATE 2-SULF	3.47e+01

21	43	75.4	389	1	034135	40 KDA HEAT SHOCK CHAP	3.47e+01
22	43	75.4	390	11	P97732	DECAY ACCELERATING FAC	3.47e+01
23	43	75.4	418	2	055223	POLYKINETIC SYNTHASE.	3.47e+01
24	43	75.4	573	10	065434	TERPENE CYCLASE LIKE P	3.47e+01
25	43	75.4	582	10	023438	SIMILAR TO.	3.47e+01
26	43	75.4	598	10	023651	TERPENE SYNTHASE, TS1.	3.47e+01
27	43	75.4	599	10	022680	TERPENE CYCLASE, TC1.	3.47e+01
28	43	75.4	602	10	064582	T29F13.4 PROTEIN.	3.47e+01
29	43	75.4	715	1	059025	715A LONG HYPOTHETICA	3.47e+01
30	43	75.4	857	2	083060	ATP-DEPENDENT PROTEASE	3.47e+01
31	43	75.4	1279	2	031827	PEPTIDE SYNTHETASE.	3.47e+01
32	43	75.4	1287	4	092549	MVELOBLAST KIAA0261 (F	3.47e+01
33	43	75.4	3121	13	042269	HUNTINGTIN.	3.47e+01
34	42	73.7	58	2	Q47054	APPA PH 2.5 ACID PHOSP	5.78e+01
35	42	73.7	201	14	Q88962	(WR) DNA-DEPENDENT RNA	5.78e+01
36	42	73.7	464	2	Q56885	HTRA GENE (FRAGMENT).	5.78e+01
37	42	73.7	478	2	P74978	GSRA PROTEIN.	5.78e+01
38	42	73.7	612	10	065437	TERPENE CYCLASE LIKE P	5.78e+01
39	42	73.7	644	5	018239	SIMILAR TO THE DISHEVE	5.78e+01
40	42	73.7	677	5	018209	Y53C12A.1 PROTEIN.	5.78e+01
41	42	73.7	700	4	015254	PRISTANOYL-COA OXIDASE	5.78e+01
42	42	73.7	727	6	Q28161	PLAKOPHILIN.	5.78e+01
43	42	73.7	728	11	P97350	PLAKOPHILIN 1.	5.78e+01
44	42	73.7	803	5	Q22343	T08D10.2 PROTEIN.	5.78e+01
45	42	73.7	1213	3	Q14340	HYPOTHETICAL 137.8 KD	5.78e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	393 AA.
ID	088371			
AC	088371			
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DE	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
GN	SEMAPHORIN L (FRAGMENT).			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 98389619.			
RA	LONGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;			
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA			
RT	viruses."			
RL	GENOMICS 51:340-350(1998).			
DR	EMBL; AF030699; G3523117; -.			
FT	NON_TER 393			
SQ	SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;			

Query Match 100.0%; Score 57; DB 11; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.15e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	205 RGESELYT 212
QY	175 RGESELYT 182

RESULT 2
ID 075326 PRELIMINARY; PRT; 666 AA.

AC	075326;
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN	SEMAPHORIN L.
OS	HOMO SAPIENS (HUMAN).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC	CATARRHINI; HOMINIDAE; HOMO.
RN	[1]
RP	SEQUENCE FROM N.A.

```
RX MEDLINE: 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL: AF030698; G3521115; -.
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 57; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 1.15e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 207 RGESELYT 214
|||
QY 175 RGESELYT 182
|||

RESULT 3
ID O51548 PRELIMINARY; PRT; 250 AA.
AC O51548;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE CHAPERONIN, PUTATIVE.
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:ATCC 35210 / B31;
RX MEDLINE: 98065943.
RA FRASER C.M., CASSENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA VAN VUOT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi."
RL NATURE 390:580-586(1997).
DR EMBL: AE001161; G2688511; -.
DR TIGR: BB0602; -.
SQ SEQUENCE 250 AA; 29473 MW; 889BBC72 CRC32;

Query Match 87.7%; Score 50; DB 2; Length 250;
Best Local Similarity 87.5%; Pred. No. 7.50e-01;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 99 RKGSELYT 106
|||
QY 175 RGESELYT 182
|||

RESULT 4
ID Q56923 PRELIMINARY; PRT; 67 AA.
AC Q56923;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN (FRAGMENT).
OS YERSINIA ENTEROCOLITICA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC YERSINIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:w22703;
RX MEDLINE: 89123026.
RA CORNELIS G., SLUITERS C., DE ROUVROIT C.L., MICHELIS T.;
RT "Homology between virF, the transcriptional activator of the Yersinia
RT virulence regulon, and AraC, the Escherichia coli arabinose operon
RT regulator.";
```

```
RL J. BACTERIOL. 171:254-262(1989).
DR EMBL: M22781; G1197046; -.
KW HYPOTHETICAL PROTEIN.
FT NON_TER 1
SQ SEQUENCE 67 AA; 7490 MW; 55D4A464 CRC32;

Query Match 86.0%; Score 49; DB 2; Length 67;
Best Local Similarity 85.7%; Pred. No. 1.35e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 16 RGESELYT 22
|||
QY 175 RGESELYT 181
|||

RESULT 5
ID Q56851 PRELIMINARY; PRT; 131 AA.
AC Q56851;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PRE-LIPOPROTEIN VIRG PRECURSOR.
OS YERSINIA ENTEROCOLITICA.
OG PLASMID PVEZ27.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC YERSINIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:w22703;
RX MEDLINE: 95362644.
RA ALLAOUI A., SCHEEN R., DE ROUVROIT C., CORNELIS G.R.;
RT "VirG, a Yersinia enterocolitica lipoprotein involved in Ca2+
RT dependency, is related to exsB of Pseudomonas aeruginosa.";
RL J. BACTERIOL. 177:4230-4237(1995).
DR EMBL: U21297; G885972; -.
KW SIGNAL; LIPOPROTEIN; PLASMID.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 131 LIPOPROTEIN VIRG.
SQ SEQUENCE 131 AA; 14687 MW; 728A0775 CRC32;

Query Match 86.0%; Score 49; DB 2; Length 131;
Best Local Similarity 85.7%; Pred. No. 1.35e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 80 RGESELYT 86
|||
QY 175 RGESELYT 181
|||

RESULT 6
ID O68694 PRELIMINARY; PRT; 131 AA.
AC O68694;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PRE-LIPOPROTEIN VIRG HOMOLOG.
GN VIRG.
OS YERSINIA PESTIS.
OG PLASMID PCD1.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC YERSINIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KIM;
RA HU P., ELLIOTT J., MCCREADY P., SKOWRONSKI E., GARNES J.,
RA KOBAYASHI A., CARRANO A.V., BRUBAKER R., GARCIA E.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF053946; G2996232; -.
KW LIPOPROTEIN; PLASMID.
SQ SEQUENCE 131 AA; 14712 MW; 490F359F CRC32;

Query Match 86.0%; Score 49; DB 2; Length 131;
Best Local Similarity 85.7%; Pred. No. 1.35e+00;
```

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 80 RGEGLY 86
|||:||||
Qy 175 RGESELY 181

RESULT 7
ID Q17233 PRELIMINARY; PRT; 467 AA.
AC Q17233;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE P50 PROTEIN PRECURSOR.
OS BOMBYX MORI (SILK MOTH).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; LEPIDOPTERA; BOMBYCOIDEA; BOMBYCIDAE; BOMBYX.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE; 96353914.
RA LEE W.J., LEE J.D., KRAVCHENKO V.V., ULEVITCH R.J., BREY P.T.;
RT "Purification and molecular cloning of an inducible gram-negative
bacteria-binding protein from the silkworm, Bombyx mori.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:7888-7893(1996).
DR EMBL; L38591; G1209253; -.
KW SIGNAL.
FT SIGNAL. 1 18 POTENTIAL.
FT CHAIN 19 467 P50 PROTEIN.
SQ SEQUENCE 467 AA; 52217 MW; 75D51BA1 CRC32;

Query Match 82.5%; Score 47; DB 5; Length 467;
Best Local Similarity 62.5%; Pred. No. 4.11e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 299 RGNALYS 306
|||:||||
Qy 175 RGESELY 182

RESULT 8
ID P96223 PRELIMINARY; PRT; 489 AA.
AC P96223;
DT 01-MAY-1997 (TREMREL. 03, CREATED)
DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 55.3 KD PROTEIN.
GN MTCY01A6.14.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA BADCOCK K., CHURCHER C.M.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
leprae.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
DR EMBL; Z83864; E301250; -.

KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 489 AA; 55326 MW; EC915838 CRC32;

Query Match 78.9%; Score 45; DB 2; Length 489;
Best Local Similarity 50.0%; Pred. No. 1.22e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 54 RSDSDMYT 61
|:|:|:|
Qy 175 RGESELYT 182

RESULT 9
ID O46307 PRELIMINARY; PRT; 771 AA.
AC O46307;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)
DE EG-8D8.4 PROTEIN.
GN EG-8D8.4.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RA PAPAGIANNAKIS G., SPANOS L., COX S., SIDEN-KIAMOS I., LOUIS C.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA BENOS P.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AL022018; E1254565; -.
SQ SEQUENCE 771 AA; 88529 MW; B2D63D51 CRC32;

Query Match 78.9%; Score 45; DB 5; Length 771;
Best Local Similarity 75.0%; Pred. No. 1.22e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 180 RGEGRLYT 187
|||:|
Qy 175 RGESELYT 182

RESULT 10
ID Q57093 PRELIMINARY; PRT; 192 AA.
AC Q57093;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE RPOE, LEPA AND ORF 2, 3, 4 GENES.
GN RPOE.
OS PHOTOBACTERIUM SP.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIONACEAE;
OC PHOTOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS9;
RA CHI E., BARTLETT D.H.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SS9;
RA CHI E., BARTLETT D.H.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; L41688; G780317; -.
DR EMBL; L41667; G777747; -.
DR PROSITE; PS01063; SIGMA70_ECF; 1.
DR PFAM; PF00776; SIGMA70_ECF; 1.
SQ SEQUENCE 192 AA; 21858 MW; 27804D3B CRC32;

Query Match 77.2%; Score 44; DB 2; Length 192;
Best Local Similarity 75.0%; Pred. No. 2.06e+01;

```
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 66 RGEAFVT 73
    ||||:|
QY 175 RGESELYT 182

RESULT 11
ID O93336 PRELIMINARY; PRT; 356 AA.
AC O93336;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE HEPARAN SULFATE 2-SULFOTRANSFERASE.
GN HS2ST.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 98301444.
RA BULLOCK S.L., FLETCHER J.M., BEDDINGTON R.S.P., WILSON V.A.;
RT "Renal agenesis in mice homozygous for a gene trap mutation in the
RT gene encoding heparan sulfate 2-sulfotransferase.";
RL GENES DEV. 12:1894-1906(1998).
DR ENBL; AF060179; G3228538; -.
KW TRANSFERASE.
SQ SEQUENCE 356 AA; 41749 MW; 4166094A CRC32;

Query Match 77.2%; Score 44; DB 13; Length 356;
Best Local Similarity 71.4%; Pred. No. 2.06e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 273 RGATELY 279
    ||:|
QY 175 RGESELYT 181

RESULT 12
ID O75036 PRELIMINARY; PRT; 356 AA.
AC O75036;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE KIAA0448 PROTEIN.
DE KIAA0448 PROTEIN.
GN KIAA0448
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RN SEQUENCE FROM N.A.
RX TISSUE-BRAIN.
RX MEDLINE; 98116662.
RA SEKI N., OHRA M., NAGASE T., ISHIKAWA K., MIYAJIMA N., NAKAJIMA D.,
RA NOMURA N., OHARA O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain.";
RL DNA RES. 4:345-349(1997).
DR ENBL; AB007917; D1033255; -.
SQ SEQUENCE 356 AA; 41881 MW; A62050E2 CRC32;

Query Match 77.2%; Score 44; DB 4; Length 356;
Best Local Similarity 71.4%; Pred. No. 2.06e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 273 RGATELY 279
    ||:|
QY 175 RGESELYT 181

RESULT 13
ID O82549 PRELIMINARY; PRT; 810 AA.
AC O82549;
```

```
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOLIPASE D1 (EC 3.1.4.4) (LIPOPHOSPHODIESTERASE II)
DE (LECITHINASE D) (CHOLINE PHOSPHATASE).
GN PLD1.
OS BRASSICA OLERACEA (CAULIFLOWER).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; BRASSICA.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. CAPITATA (LENNOX);
RA PANNENBERG I., MANSFELD J., ULBRICH-HOFMANN R.;
RT "Cloning of phospholipase D from Brassica oleracea var. capitata.";
RT SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -|- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O -> CHOLINE + A
CC PHOSPHATIDATE.
DR ENBL; AF090445; G3639089; -.
KW HYDROLASE.
SQ SEQUENCE 810 AA; 91837 MW; 3F0F8FC3 CRC32;

Query Match 77.2%; Score 44; DB 10; Length 810;
Best Local Similarity 50.0%; Pred. No. 2.06e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 45 KGETOLYA 52
    ||:|
QY 175 RGESELYT 182

RESULT 14
ID O49981 PRELIMINARY; PRT; 812 AA.
AC O49981;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 09, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE PHOSPHOLIPASE D (EC 3.1.4.4) (LIPOPHOSPHODIESTERASE II) (LECITHINASE
DE D) (CHOLINE PHOSPHATASE).
OS BRASSICA OLERACEA VAR. CAPITATA (CABBAGE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; BRASSICA.
RN [1]
RN SEQUENCE FROM N.A.
RA KIM D.U., LEE J.E., HOE K.R., JANG Y.J., RHO T.Y., YOO H.S.,
RA CHOI M.-U.;
RT "Cloning of phospholipase D from cabbage.";
RL SUBMITTED (NOV-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -|- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O -> CHOLINE + A
CC PHOSPHATIDATE.
DR ENBL; U85482; G3924621; -.
KW HYDROLASE.
SQ SEQUENCE 812 AA; 92062 MW; 4F3BD14F CRC32;

Query Match 77.2%; Score 44; DB 10; Length 812;
Best Local Similarity 50.0%; Pred. No. 2.06e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 46 KGETOLYA 53
    ||:|
QY 175 RGESELYT 182

RESULT 15
ID O82548 PRELIMINARY; PRT; 812 AA.
AC O82548;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOLIPASE D2 (EC 3.1.4.4) (LIPOPHOSPHODIESTERASE II)
DE (LECITHINASE D) (CHOLINE PHOSPHATASE).
GN PLD2.
```

OS BRASSICA OLERACEA (CAULIFLOWER).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; BRASSICA.

[1]

RN SEQUENCE FROM N.A.
 RC STRAIN=CV. CAPITATA (LENNOX);
 RA PANNENBERG I., MANSFELD J., ULBRICH-HOFMANN R.;
 RT "Cloning of phospholipase D from Brassica oleracea var. capitata.";
 RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
 CC -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
 CC PHOSPHATIDATE.
 DR EMBL; AF090444; G3639087; -.
 KW HYDROLASE.
 SQ SEQUENCE 812 AA; 92062 MW; 4F3BD14F CRC32; -

Query Match 77.2%; Score 44; DB 10; Length 812;
 Best Local Similarity 50.0%; Pred. NO. 2.06e+01;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 46 KGETOLYA 53

Oy 175 RGESELYT 182

Search completed: Thu Jul 8 18:25:29 1999
 Job time : 13 secs.

THIS PAGE BLANK (USPTO)

WQSRH

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:28:32 1999; MasPar time 6.88 Seconds
49.445 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (180-195) from US09041236.pep (11 of 45)
Perfect Score: 114
Sequence: 1 LYTSDTVMQNPFKA 16

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-genes35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 19.098; Variance 59.936; scale 0.319

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	59	51.8	921	20	W11252	3.40e+01
2	59	51.8	921	5	R28407	3.40e+01
3	59	51.8	921	13	R69727	3.40e+01
4	59	51.8	921	22	W18048	3.40e+01
5	59	51.8	921	32	W60752	3.40e+01
6	59	51.8	921	34	W11224	3.40e+01
7	59	51.8	921	35	W77040	3.40e+01
8	59	51.8	921	22	W18050	3.40e+01
9	59	51.8	921	34	W11253	3.40e+01
10	59	51.8	921	34	W71226	3.40e+01
11	59	51.8	921	5	R28409	3.40e+01
12	59	51.8	921	13	R69729	3.40e+01
13	59	51.8	921	32	W60754	3.40e+01
14	59	51.8	921	36	W77042	3.40e+01
15	59	51.8	922	32	W60753	3.40e+01
16	59	51.8	942	13	R69728	3.40e+01

17	59	51.8	942	35	W77041	Cyclic-GMP-stimulated	3.40e+01
18	59	51.8	942	22	W18049	CGS-PDE amino acid se	3.40e+01
19	59	51.8	942	5	R28408	CGS-PDE from bovine b	3.40e+01
20	59	51.8	942	20	W11239	Cyclic GMP stimulated	3.40e+01
21	59	51.8	942	34	W71225	Bovine brain CGS-PDE	3.40e+01
22	58	50.9	378	29	W55204	H. pylori ORF 01cel16	4.31e+01
23	58	50.9	431	29	W55553	H. pylori ORF 02ae310	4.31e+01
24	58	50.9	431	30	W55533	H. pylori ORF 02ae310	4.31e+01
25	58	50.9	510	37	W79741	Soybean mutant myo-in	4.31e+01
26	58	50.9	510	37	W79740	Soybean wild-type myo	4.31e+01
27	56	49.1	1150	33	W62541	Bacillus thuringiens	6.90e+01
28	56	49.1	1150	29	W46858	Bacillus thuringiens	6.90e+01
29	55	48.2	109	21	W09792	Lectin active protein	8.71e+01
30	55	48.2	409	34	W71495	Helicobacter polypept	8.71e+01
31	55	48.2	504	14	R72509	Hamster cholesterol 7	8.71e+01
32	54	47.4	187	38	R83903	Porcine interferon-be	1.10e+02
33	54	47.4	187	38	W73232	Porcine interferon-be	1.10e+02
34	54	47.4	371	23	W23656	EGAP-binding protein	1.10e+02
35	54	47.4	377	25	W24558	Presenilin-Interactin	1.10e+02
36	54	47.4	382	21	W12829	Human recombinant ant	1.10e+02
37	54	47.4	384	28	W27303	Human prostate protei	1.10e+02
38	54	47.4	644	21	W13821	Yeast transcription r	1.10e+02
39	53	46.5	218	3	R20031	Rat liver glutathione	1.38e+02
40	53	46.5	218	1	P82925	Glutathione S-transfe	1.38e+02
41	53	46.5	300	11	R60603	Yellowjacket phosphol	1.38e+02
42	53	46.5	447	35	W70402	Anabaena biotin bioti	1.38e+02
43	53	46.5	447	20	W05207	Anabaena biotin bioti	1.38e+02
44	53	46.5	448	10	R51080	Biotin carboxylase (f	1.38e+02
45	53	46.5	502	3	R15226	HincII restriction en	1.38e+02

ALIGNMENTS

RESULT 1
ID W11252 standard; Protein: 921 AA.

AC W11252;
DT 17-MAR-1997 (first entry)
DE Clone p3CGS-5 cyclic GMP stimulated phosphodiesterase.
KW Cyclic GMP stimulated phosphodiesterase; CGS-PDE; bovine; brain; CAMP;
KW adrenal gland; cGMP; transmembrane signal; extracellular hormone;
KW neurotransmitter; antibody.
OS Bos taurus
PN US5580771-A.
PD 03-DEC-1996.
PF 19-APR-1991; 688356.
PR 19-APR-1991; US-688356.
PR 20-APR-1992; US-872644.
PR 29-AUG-1994; US-297494.
PA (UNITW) UNIV WASHINGTON.
PI Beavo JA, Charbonneau H, Sonnenburg WK;
DR WPI; 97-033573/03.
DR N-PSDB; T51110.
PT DNA encoding cGMP-stimulated phosphodiesterase - for prodn. of
PT recombinant enzyme
PS Example 4; Column 65-72; 68pp; English.
CC This sequence represents the cyclic GMP stimulated phosphodiesterase
CC (CGS-PDE) clone p3CGS-5 isolated from bovine adrenal cortex. The cyclic
CC nucleotide phosphodiesterases (PDEs) catalyze the hydrolysis of 3', 5',
CC cyclic nucleotides, such as cAMP and cGMP, to their corresponding
CC 5'-nucleotide monophosphates. The PDEs are therefore important in the
CC control of the cellular concentration of cyclic nucleotides. The PDEs
CC are, in turn, regulated by transmembrane signals or second messenger
CC ligands such as calcium ion or cGMP. The PDEs therefore have a central
CC role in regulating the flow of information from extracellular hormones,
CC neurotransmitters, or other signals that use the cyclic nucleotides as
CC messengers. PDEs are present in most of the cells and tissues of
CC eukaryotic organisms, but only in trace amounts. CGS-PDEs are stimulated
CC by cGMP in effecting cAMP hydrolysis. This is thought to occur through a
CC noncatalytic cGMP-specific site present on the CGS-PDE. The cDNA
CC encoding this sequence can be used for the production of recombinant
CC CGS-PDE, which may have therapeutic and diagnostic uses. This protein
CC may also be useful for diagnostic antibody production.
SQ Sequence 921 AA;

```

CC nucleotide phosphodiesterase (Cam PDE) clone p3CGS-5. Eukaryotic cells
CC that express Cam PDE can be used to screen cpds. for the ability to
CC modulate Cam PDE activity. Cam PDEs are involved in regulating
CC the transmission of information from hormones, neurotransmitters
CC or other systems that use cyclic nucleotides as messengers.
CC Antibodies raised against Cam PDE can be used for enzyme purifcn.,
CC or determination.
SQ Sequence 921 AA;

Query Match 51.8%; Score 59; DB 13; Length 921;
Best Local Similarity 41.2%; Pred. No. 3.40e+01;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 698 lyssegsvmerhhfaqa 714
|||: |||: |
QY 180 LYTSD-TVMQNPQFIKA 195

RESULT 4
ID W18048 standard; Protein; 921 AA.
AC W18048:
DT 05-AUG-1997 (first entry)
DE Cyclic-GMP-stimulated phosphodiesterase (cGS-PDE) from p3CGS-5.
KW Ca2+/calmodulin stimulated cyclic nucleotide phosphodiesterase enzyme;
KW calcium ion; antibody; cow; polymerase chain reaction.
OS Bos taurus.
PN US5602019-A.
PD 11-FEB-1997. 688356.
PF 19-APR-1991; US-688356.
PR 19-APR-1991; US-688356.
PR 20-APR-1992; US-872644.
PR 29-AUG-1994; US-297510.
PA (UNIW ) UNIV WASHINGTON.
PI Beavo JA, Bentley KJ, Charbonneau H, Sonnenburg WK;
DR WPI; 97-131799/12.
DR N-PSDB; T67220.
PT DNA encoding bovine and human phosphodiesterase enzymes - stimulated
PT by calcium/calmodulin, useful for recombinant prodn. of the enzymes
PS Example 4; Column 73-78; 69pp; English.
CC The present sequence represents the amino acid sequence from the large
CC open reading frame of p3CGS-5 cDNA bovine adrenal cortex clone insert.
CC The presence of 36 adenosine residues at the 3'-end of the cDNA preceded
CC by a transcription termination consensus sequence suggests that all of
CC the 3' untranslated sequence of the cyclic-GMP-stimulated
CC phosphodiesterase (cGS-PDE) mRNA is represented by this clone. The DNA
CC sequences of the cGS-PDE's are used for the production of the
CC recombinant enzymes, which in turn may be used for antibody production
CC and to screen for compounds that modulate phosphodiesterase activity.
SQ Sequence 921 AA;

Query Match 51.8%; Score 59; DB 22; Length 921;
Best Local Similarity 41.2%; Pred. No. 3.40e+01;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 698 lyssegsvmerhhfaqa 714
|||: |||: |
QY 180 LYTSD-TVMQNPQFIKA 195

RESULT 5
ID W60752 standard; Protein; 921 AA.
AC W60752:
DT 07-SEP-1998 (first entry)
DE cGS-PDE isolated from bovine adrenal cortex.
KW Bovine; decrease; intracellular concentration; cAMP; cGMP;
KW calcium ion/calmodulin stimulated cyclic nucleotide phosphodiesterase;
KW Cam-PDE; therapy; prognosis; diagnosis; specific modulator;
KW Cam-PDE activity; identification; signalling; extracellular hormone;
KW neurotransmitter; selective inhibitor; cardiotoxic; antidiuretic;
KW antihypertensive; antithrombotic agent;
KW cyclic-GMP-nucleotide phosphodiesterase; cGS-PDE.
OS Bos sp.
PN US5776752-A.

```

07-JUL-1998.
PF 07-JUN-1995; 479532.
PR 20-APR-1992; US-872644.
PR 19-APR-1991; US-688356.
PR 29-AUG-1994; US-297494.
PA (UNIW) UNIV WASHINGTON.
PI Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;
DR WPI; 98-398038/34.
DR N-PSDB; V36153.
PT New calmodulin and calcium stimulated cyclic nucleotide
phosphodiesterase - used e.g to identify selective modulators,
potentially useful as e.g. cardiotonic, antidepressant,
antihypertensive agents
PS Example 4; Columns 71-76; 69pp; English.
CC The present sequence represents a cyclic-GMP-nucleotide phosphodiesterase
(CGS-PDE) polypeptide isolated from bovine adrenal cortex.
CC The specification describes calcium ion/calmodulin stimulated cyclic
nucleotide phosphodiesterases (CAM-PDEs). CAM-PDEs are characterised
by their responsiveness to intracellular calcium, which leads to a
decreased intracellular concentration of cAMP and/or cGMP. CAM-PDEs,
and their fragments or synthetic peptides, are used in therapy,
prognosis and diagnosis. They are also used to generate specific
antibodies useful as immunoassay reagents for detecting the
protein or for affinity purification and for identifying specific
modulators of CAM-PDE activity. CAM-PDEs are implicated in signalling
involving extracellular hormones and neurotransmitters, so selective
inhibitors are potentially useful as cardiotonic, antidepressant,
antihypertensive and antithrombotic agents.
SQ Sequence 921 AA;

Query Match 51.8%; Score 59; DB 32; Length 921;
Best Local Similarity 41.2%; Pred. No. 3.40e+01;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 698 lyssegsvmerhhfaqa 714
|||: |||: |||
Qy 180 LYTSD-TVMQNPQFIKA 195

RESULT 6
ID W71224 standard; Protein; 921 AA.
AC W71224;
DT 29-OCT-1998 (first entry)
DE CGS-PDE encoded by p3CGS-5.
KW Ca2+/calmodulin sensitive cyclic nucleotide phosphodiesterase; CAM-PDE;
KW assay; identification; modification; enzymatic activity; modulator;
KW cyclic guanosine monophosphate-stimulated phosphodiesterase; CGS-PDE.
OS Bos sp.
PN US5800987-A.
PD 01-SEP-1998.
PF 31-MAY-1995; 455525.
PR 20-APR-1992; US-872644.
PR 19-APR-1991; US-688356.
PR 29-AUG-1994; US-297494.
PR 31-MAY-1995; US-455525.
PA (UNIW) UNIV WASHINGTON.
PI Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;
DR WPI; 98-494762/42.
DR N-PSDB; V54758.
PT Identification of modulators of Calcium/calmodulin sensitive cyclic
nucleotide phosphodiesterase - uses recombinant cells expressing the
enzyme and monitoring protein expression in the presence of
potential modulators
PS Example 4; Columns 73-78; 69pp; English.
CC The present sequence is encoded by a sequence isolated from a bovine
adrenal cortex cDNA library known as cyclic guanosine monophosphate
stimulated phosphodiesterase (CGS-PDE). The specification also describes
Ca2+/calmodulin sensitive cyclic nucleotide phosphodiesterase (CAM-PDE)
enzymes. CAM-PDE enzymes catalyse the hydrolysis of cyclic nucleotides
and as such control their intracellular level. They are controlled by
second messengers e.g. Ca2+ and calmodulin and transmembrane signals and
through pathways involving these, regulate flow of information from
extracellular hormones, neurotransmitters and other signals using cyclic

CC nucleotides as messengers. The specification describes an assay for
CC identifying a chemical agent which modifies the enzymatic activity of a
CC mammalian CAM-PDE. The assays are useful for identifying modulators of
CC CAM-PDEs. The assay is carried out by incubating cells expressing
CC CAM-PDE with the suspected modulator and measuring its effect
CC e.g. monitoring the hydrolysis of cAMP and/or cGMP.
SQ Sequence 921 AA;

Query Match 51.8%; Score 59; DB 34; Length 921;
Best Local Similarity 41.2%; Pred. No. 3.40e+01;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 698 lyssegsvmerhhfaqa 714
|||: |||: |||
Qy 180 LYTSD-TVMQNPQFIKA 195

RESULT 7
ID W77040 standard; Protein; 921 AA.
AC W77040;
DT 09-NOV-1998 (first entry)
DE Adrenal cortex Ca2+/calmodulin stimulated phosphodiesterase.
DE ss; cow; Ca2+ stimulated phosphodiesterase;
KW Calmodulin stimulated phosphodiesterase; therapeutic; antibody.
OS Bos sp.
PN US578553-A.
PD 04-AUG-1998.
PF 31-MAY-1995; 455526.
PR 20-APR-1992; US-872644.
PR 19-APR-1991; US-688356.
PR 29-AUG-1994; US-297494.
PR 31-MAY-1995; US-455526.
PA (UNIW) UNIV WASHINGTON.
PI Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;
DR WPI; 98-446185/38.
DR N-PSDB; V48201.
PT Antibody to phosphodiesterase polypeptide - useful for purification
PT or detection of polypeptide
PS Example 4; Column 65-72; 68pp; English.
CC The adrenal cortex Ca2+/calmodulin stimulated phosphodiesterase is one of
CC several phosphodiesterases isolated in this invention, used to transform
CC host cells, preferably mammalian or yeast cells. The recombinant
CC proteins and fragments produced can be used for therapeutic, diagnostic,
CC and prognostic purposes and will provide the basis for preparation of
CC monoclonal and polyclonal antibodies. An antibody specific for the
CC Ca2+/calmodulin-stimulated cyclic nucleotide phosphodiesterase (PDE)
CC polypeptide is useful for affinity purification or detection of the
CC polypeptide and is selective only for the Ca2+/calmodulin stimulated
CC cyclic nucleotide PDE polypeptide.
SQ Sequence 921 AA;

Query Match 51.8%; Score 59; DB 35; Length 921;
Best Local Similarity 41.2%; Pred. No. 3.40e+01;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 698 lyssegsvmerhhfaqa 714
|||: |||: |||
Qy 180 LYTSD-TVMQNPQFIKA 195

RESULT 8
ID W18050 standard; Protein; 941 AA.
AC W18050;
DT 05-AUG-1997 (first entry)
DE Human CGS-PDE amino acid sequence from pHcgs6n.
KW Cyclic-GMP-stimulated phosphodiesterase; antibody; human;
KW polymerase chain reaction.
OS Homo sapiens.
PN US5602019-A.
PD 11-FEB-1997.
PF 19-APR-1991; 688356.
PR 19-APR-1991; US-688356.
PR 20-APR-1992; US-872644.

PR 29-AUG-1994; US-297510.
PA (UNIW) UNIV WASHINGTON.
PI Beavo JA, Bentley KJ, Charbonneau H, Sonnenburg WK;
DR WPI: 97-131799/12.
DR N-PSDB; T67224.
PT DNA encoding bovine and human phosphodiesterase enzymes - stimulated
PT by calcium/calmodulin, useful for recombinant prodn. of the enzymes
PS Example 6; Column 97-102; 69pp; English.
CC The present sequence represents the human cyclic-GMP-stimulated
CC phosphodiesterase (cGS-PDE) from pHCs6n plasmid. The DNA sequences
CC of the cGS-PDE's are used for the production of the recombinant
CC enzymes, which in turn may be used for antibody production and to
CC screen for compounds that modulate phosphodiesterase activity.
SQ Sequence 941 AA;

Query Match 51.8%; Score 59; DB 22; Length 941;
Best Local Similarity 41.2%; Pred. No. 3.40e+01;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 718 lyssegsvmerhhfaga 734
||:|: |||: | |
QY 180 LYTSD-TVMQNPOFIKA 195

RESULT 9
ID W11253 standard; Protein: 941 AA.
AC W11253;
DT 17-MAR-1997 (first entry)
DE pHCs6n cyclic GMP stimulated phosphodiesterase.
KW Cyclic GMP stimulated phosphodiesterase; cGS-PDE; bovine; brain; cAMP;
KW adrenal gland; cGMP; transmembrane signal; extracellular hormone;
KW neurotransmitter; antibody; human.
OS Homo sapiens.
PN US5580771-A.
PD 03-DEC-1996.
PF 19-APR-1991; 688356.
PR 19-APR-1991; US-688356.
PR 20-APR-1992; US-872644.
PR 29-AUG-1994; US-297494.
PA (UNIW) UNIV WASHINGTON.
PI Beavo JA, Charbonneau H, Sonnenburg WK;
DR WPI: 97-033573/03.
DR N-PSDB; T51113.
PT DNA encoding cGMP-stimulated phospho-di:esterase - for prodn. of
PT recombinant enzyme
PS Example 6; Column 89-96; 68pp; English.
CC This sequence represents the cyclic GMP stimulated phosphodiesterase
CC (cGS-PDE) clone pHCs6n isolated from human foetal brain. The cyclic
CC nucleotide phosphodiesterases (PDEs) catalyse the hydrolysis of 3', 5'
CC cyclic nucleotides, such as cAMP and cGMP, to their corresponding
CC 5'-nucleotide monophosphates. The PDEs are therefore important in the
CC control of the cellular concentration of cyclic nucleotides. The PDEs
CC are, in turn, regulated by transmembrane signals or second messenger
CC ligands such as calcium ion or cGMP. The PDEs therefore have a central
CC role in regulating the flow of information from extracellular hormones,
CC neurotransmitters, or other signals that use the cyclic nucleotides as
CC messengers. PDEs are present in most of the cells and tissues of
CC eukaryotic organisms, but only in trace amounts. cGS-PDEs are stimulated
CC by cGMP in effecting cAMP hydrolysis. This is thought to occur through a
CC noncatalytic cGMP-specific site present on the cGS-PDE. The cDNA
CC encoding this sequence can be used for the production of recombinant
CC cGS-PDE, which may have therapeutic and diagnostic uses. This protein
CC may also be useful for diagnostic antibody production.
SQ Sequence 941 AA;

Query Match 51.8%; Score 59; DB 20; Length 941;
Best Local Similarity 41.2%; Pred. No. 3.40e+01;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 718 lyssegsvmerhhfaga 734
||:|: |||: | |
QY 180 LYTSD-TVMQNPOFIKA 195

RESULT 10
ID W71226 standard; Protein: 941 AA.
AC W71226;
DT 29-OCT-1998 (first entry)
DE Human cGS-PDE cDNA protein sequence.
KW Ca2+/calmodulin sensitive cyclic nucleotide phosphodiesterase; Cam-PDE;
KW assay; identification; modification; enzymatic activity; modulator;
KW cyclic guanosine monophosphate-stimulated phosphodiesterase; cGS-PDE;
KW human.
OS Homo sapiens.
PN US5800987-A.
PD 01-SEP-1998.
PF 31-MAY-1995; 455525.
PR 20-APR-1992; US-872644.
PR 19-APR-1991; US-688356.
PR 29-AUG-1994; US-297494.
PR 31-MAY-1995; US-455525.
PA (UNIW) UNIV WASHINGTON.
PI Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;
DR WPI: 98-494762/42.
DR N-PSDB; V54762.
PT Identification of modulators of Calcium/calmodulin sensitive cyclic
PT nucleotide phosphodiesterase - uses recombinant cells expressing the
PT enzyme and monitoring protein expression in the presence of
PT potential modulators
PS Example 6; Columns 97-102; 69pp; English.
CC The present sequence represents human cyclic guanosine
CC monophosphate-stimulated phosphodiesterase (cGS-PDE) cDNA. The
CC specification also describes Ca2+/calmodulin sensitive cyclic
CC nucleotide phosphodiesterase (Cam-PDE) enzymes. Cam-PDE enzymes
CC catalyse the hydrolysis of cyclic nucleotides and as such control their
CC intracellular level. They are controlled by second messengers e.g. Ca2+
CC and calmodulin and transmembrane signals and through pathways involving
CC these, regulate flow of information from extracellular hormones,
CC neurotransmitters and other signals using cyclic nucleotides as
CC messengers. The specification describes an assay for identifying a
CC chemical agent which modifies the enzymatic activity of a mammalian
CC Cam-PDE. The assays are useful for identifying modulators of
CC Cam-PDEs. The assay is carried out by incubating cells expressing
CC Cam-PDE with the suspected modulator and measuring its effect
CC e.g. monitoring the hydrolysis of cAMP and/or cGMP.
SQ Sequence 941 AA;

Query Match 51.8%; Score 59; DB 34; Length 941;
Best Local Similarity 41.2%; Pred. No. 3.40e+01;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 718 lyssegsvmerhhfaga 734
||:|: |||: | |
QY 180 LYTSD-TVMQNPOFIKA 195

RESULT 11
ID R28409 standard; Protein: 941 AA.
AC R28409;
DT 19-MAR-1993 (first entry)
DE Human foetal cGS PDE from clone HFB9.2.
KW Cyclic GMP; stimulated; cyclic; nucleotide; bovine brain;
KW phosphodiesterase; cGS PDE; hybridisation.
OS Homo sapiens.
PN WO9218541-A.
PD 29-OCT-1992.
PF 20-APR-1992; U03222.
PR 19-APR-1991; US-688356.
PA (UNIW) UNIV WASHINGTON.
PI Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;
DR WPI: 92-382051/46.
DR N-PSDB; Q30184.
PT New DNA encoding mammalian cyclic nucleotide phospho-di:esterase
PT - and derived vectors and host cells, useful for screening cpds.
PT for inhibitory or activating activity
PS Example 6; Page 98; 133pp; English.

CC An approx. 0.95 kb EcoRI-HindII fragment from the 5' end of the
 CC bovine cyclic GMP stimulated phosphodiesterase cDNA plasmid p3CGS-5
 CC was random primed and used as a probe to screen a human foetal brain
 CC cDNA library, revealing two positive clones HFB9.1 and HFB9.2. HFB9.2
 CC is the longer clone and covers the entire ORF of cgs pDE but reads
 CC into what may be an intron 59 nucleotides after the stop codon.
 CC The clone lacks the 5' 0.4 kb and presumed intron found in pGSPDE9.2.
 CC The entire ORF of HFB9.2 was isolated and inserted into the yeast
 CC expression vector pANY6N.
 CC Sequence 941 AA;

Query Match 51.8%; Score 59; DB 5; Length 941;
 Best Local Similarity 41.2%; Pred. No. 3.40e+01;
 Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 718 lyssegsvmerhhfaga 734
 |||: |||: |||
 Qy 180 LYTSD-TVMQNPOFIKA 195

RESULT 12

ID R69729 standard; Protein: 941 AA.

AC R69729;
 DT 11-OCT-1995 (first entry)
 DE Cyclic-GMP stimulated nucleotide PDE clone HFB9.2.
 KW Cyclic-GMP stimulated nucleotide phosphodiesterase; antibodies;
 KW human; hormones; neurotransmitters; transmission regulation;
 KW enzyme purification; clone HFB9.2.

OS Homo sapiens.

PN US389527-A.

PD 14-FEB-1995.

PF 19-APR-1991; 688356.

PR 19-APR-1991; US-688356.

PR 20-APR-1992; US-872644.

PA (UNIW) UNIV WASHINGTON.

PI Beavo JA, Charbonneau H, Sonnenburg WK;

DR WPI: 95-090205/12.

DR N-PSDB: 083978.

PT New nucleic acid encoding cyclic-GMP stimulated nucleotide
 PT phospho-diesterase - and related vectors and transformed cells,
 PT useful for screening cpds. for phospho-diesterase modulating
 PT activity

PS Claim 1; Columns 89-96; 69pp; English.

CC 083978 encodes R69729 the human cyclic-GMP stimulated nucleotide
 CC phosphodiesterase (CaM PDE) clone HFB9.2. Eukaryotic cells
 CC that express CaM PDE can be used to screen cpds. for the ability to
 CC modulate CaM PDE activity. CaM PDEs are involved in regulating
 CC the transmission of information from hormones, neurotransmitters
 CC or other systems that use cyclic nucleotides as messengers.
 CC Antibodies raised against CaM PDE can be used for enzyme purific.,
 CC or determination.

CC Sequence 941 AA;

Query Match 51.8%; Score 59; DB 13; Length 941;
 Best Local Similarity 41.2%; Pred. No. 3.40e+01;
 Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 718 lyssegsvmerhhfaga 734
 |||: |||: |||
 Qy 180 LYTSD-TVMQNPOFIKA 195

RESULT 13

ID W60754 standard; Protein: 941 AA.

AC W60754;

DT 07-SEP-1998 (first entry)

DE Human cyclic-GMP-nucleotide phosphodiesterase.

KW Bovine; decrease; intracellular concentration; cAMP; CGMP;

KW calcium ion/calmodulin stimulated cyclic nucleotide phosphodiesterase;

KW CaM-PDE; therapy; prognosis; diagnosis; specific modulator;

KW CaM-PDE activity; identification; signalling; extracellular hormone;

KW neurotransmitter; selective inhibitor; cardiotonic; antidepressant;

KW antihypertensive; antithrombotic agent;

KW cyclic-GMP-nucleotide phosphodiesterase; cGS-PDE.
 OS Homo sapiens.

PN US5776752-A.

PD 07-JUL-1998.

PF 07-JUN-1995; 479532.

PR 20-APR-1992; US-872644.

PR 19-APR-1991; US-688356.

PR 29-AUG-1994; US-297494.

PA (UNIW) UNIV WASHINGTON.

PI Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;

DR WPI: 98-398038/34.

DR N-PSDB: V36157.

PT New calmodulin and calcium stimulated cyclic nucleotide

PT phosphodiesterase - used e.g to identify selective modulators,

PT potentially useful as e.g. cardiotonic, antidepressant,

PT antihypertensive agents

PS Example 6; Columns 89-102; 69pp; English.

CC The present sequence represents human cyclic-GMP-nucleotide

CC phosphodiesterase (cGS-PDE). It was isolated using bovine adrenal cortex

CC cGS-PDEs. The specification describes calcium ion/calmodulin stimulated

CC cyclic nucleotide phosphodiesterases (CaM-PDEs). CaM-PDEs are

CC characterised by their responsiveness to intracellular calcium, which

CC leads to a decreased intracellular concentration of cAMP and/or cGMP.

CC CaM-PDEs, and their fragments or synthetic peptides, are used in therapy,

CC prognosis and diagnosis. They are also used to generate specific

CC antibodies useful as immunoassay reagents for detecting the protein or

CC for affinity purification and for identifying specific modulators of

CC CaM-PDE activity. CaM-PDEs are implicated in signalling involving

CC extracellular hormones and neurotransmitters, so selective inhibitors are

CC potentially useful as cardiotonic, antidepressant, antihypertensive and

CC antithrombotic agents.

CC Sequence 941 AA;

Query Match 51.8%; Score 59; DB 32; Length 941;

Best Local Similarity 41.2%; Pred. No. 3.40e+01;

Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 718 lyssegsvmerhhfaga 734

|||: |||: |||

Qy 180 LYTSD-TVMQNPOFIKA 195

RESULT 14

ID W77042 standard; Protein: 941 AA.

AC W77042;

DT 14-DEC-1998 (first entry)

DE Cyclic-GMP-stimulated phosphodiesterase clone pHCgs6n.

KW cow; Ca2+ stimulated phosphodiesterase;

KW Calmodulin stimulated phosphodiesterase; therapeutic; antibody.

OS Bos sp.

PN US5789553-A.

PD 04-AUG-1998.

PF 31-MAY-1995; 455526.

PR 20-APR-1992; US-872644.

PR 19-APR-1991; US-688356.

PR 29-AUG-1994; US-297494.

PR 31-MAY-1995; US-455526.

PA (UNIW) UNIV WASHINGTON.

PI Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;

DR WPI: 98-446185/38.

DR N-PSDB: V48205.

PT Antibody to phosphodiesterase polypeptide - useful for purification

PT or detection of polypeptide

PS Example 6; Column 89/96; 68pp; English.

CC The cyclic-GMP-stimulated phosphodiesterase clone pHCgs6n encodes a

CC phosphodiesterase from the bovine brain. The phosphodiesterases isolated

CC in this invention were used to transform host cells, preferably mammalian

CC or yeast cells. The recombinant proteins and fragments produced can be

CC used for therapeutic, diagnostic, and prognostic purposes and will

CC provide the basis for preparation of monoclonal and polyclonal

CC antibodies. An antibody specific for the Ca2+/calmodulin-stimulated

CC cyclic nucleotide phosphodiesterase (PDE) polypeptide is useful for

CC affinity purification or detection of the polypeptide and is selective

CC only for the Ca2+/calmodulin stimulated cyclic nucleotide PDE
CC polypeptide.
SQ Sequence 941 AA;

Query Match 51.8%; Score 59; DB 36; Length 941;
Best Local Similarity 41.2%; Pred. No. 3.40e+01;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 718 lyssegsvmerhhfaqa 734

||||: |||: ||

QY 180 LYTSD-IVMQNPQFIKA 195

RESULT 15

ID W60753 standard; Protein: 942 AA.

AC W60753;

DT 07-SEP-1998 (first entry)

DE CGS-PDE isolated from bovine brain.

KW Bovine; decrease; intracellular concentration; cAMP; cGMP;

KW calcium ion/calmodulin stimulated cyclic nucleotide phosphodiesterase;

KW Cam-PDE; therapy; prognosis; diagnosis; specific modulator;

KW Cam-PDE activity; identification; signalling; extracellular hormone;

KW neurotransmitter; selective inhibitor; cardiotonic; antidepressant;

KW antihypertensive; antithrombotic agent;

KW cyclic-GMP-nucleotide phosphodiesterase; cGS-PDE.

OS Bos sp.

PN US5776752-A.

PD 07-JUL-1998.

PF 07-JUN-1995; 479532.

PR 20-APR-1992; US-872644.

PR 19-APR-1991; US-688356.

PR 29-AUG-1994; US-297494.

PA (UNIW) UNIV WASHINGTON.

PI Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;

DR WPI: 98-398038/34.

DR N-PSDB: V36156.

PT New calmodulin and calcium stimulated cyclic nucleotide

PT phosphodiesterase - used e.g to identify selective modulators,

PT potentially useful as e.g. cardiotonic, antidepressant,

PT antihypertensive agents

PS Example 5: Columns 85-90; 69pp; English.

CC The present sequence represents cyclic-GMP-nucleotide phosphodiesterase

CC (cGS-PDE), and is isolated from bovine. The specification describes

CC calcium ion/calmodulin stimulated cyclic nucleotide phosphodiesterases

CC (Cam-PDEs). Cam-PDEs are characterised by their responsiveness to

CC intracellular calcium, which leads to a decreased intracellular

CC concentration of cAMP and/or cGMP. Cam-PDEs, and their fragments or

CC synthetic peptides, are used in therapy, prognosis and diagnosis. They

CC are also used to generate specific antibodies useful as immunoassay

CC reagents for detecting the protein or for affinity purification and for

CC identifying specific modulators of Cam-PDE activity. Cam-PDEs are

CC implicated in signalling involving extracellular hormones and

CC neurotransmitters, so selective inhibitors are potentially useful as

CC cardiotonic, antidepressant, antihypertensive and antithrombotic agents.

SQ Sequence 942 AA;

Query Match 51.8%; Score 59; DB 32; Length 942;

Best Local Similarity 41.2%; Pred. No. 3.40e+01;

Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 719 lyssegsvmerhhfaqa 735

||||: |||: ||

QY 180 LYTSD-IVMQNPQFIKA 195

Search completed: Thu Jul 8 18:28:54 1999

Job time : 22 secs.

Result No.	Query			DB	ID	Description	Pred. No.
	No.	Score	Match				
1	89	78.1	653	2	T03102	semaphorin homolog A3	3.55e+06
2	62	54.4	428	2	G1023	translation elongatio	1.18e+00
3	62	54.4	428	2	S10248	translation elongatio	1.18e+00
4	62	54.4	403	2	S19000	translation elongatio	1.18e+00
5	61	53.5	203	2	S43222	hypothetical protein	1.80e+00
6	60	52.6	389	2	T00697	probable glycylpeptid	2.73e+00
7	59	51.8	115	2	A49810	hypothetical protein	4.13e+00
8	59	51.8	457	2	T02140	hypothetical protein	4.13e+00
9	59	51.8	611	2	A60179	3',5'-cyclic-nucleoti	4.13e+00
10	59	51.8	818	2	JC4397	peroxinectin precurs	4.13e+00
11	59	51.8	921	2	A40981	3',5'-cyclic-nucleoti	4.13e+00
12	59	51.8	928	2	JC2486	3',5'-cyclic-nucleoti	4.13e+00
13	58	50.9	430	2	C64554	ATP-dependent nucleas	6.22e+00
14	58	50.9	431	2	C71954	hypothetical protein	6.22e+00
15	58	50.9	1698	2	S18669	probable membrane pro	6.22e+00
16	57	50.0	945	2	A64714	helicase - Helicobact	9.31e+00
17	57	50.0	946	2	A71805	probable ATP-dependen	9.31e+00
18	57	50.0	1031	1	A38713	kinesin heavy chain -	9.31e+00
19	56	49.1	139	2	PC4217	hypothetical 139 prot	1.39e+01
20	56	49.1	393	1	A0SPAP	fructose-bisphosphate	1.39e+01
21	56	49.1	704	2	H71523	hypothetical protein	1.39e+01
22	56	49.1	1244	2	S25327	cytoskeleton assembly	1.39e+01
23	55	48.2	145	1	H8BOBB	hemoglobin beta-A,cha	2.06e+01

```
#cross-references MUID:98344137
#accession G71023
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-428 #label KAW
#cross-references GB:AP000006; NID:g3236133; PID:d1031534; PID:g3257908
#experimental_source strain OT3
#note this accession replaces an interim accession for a
sequence replaced by GenBank

GENETICS
#gene PH1484
CLASSIFICATION #superfamily translation elongation factor Tu; translation
elongation factor Tu homology
FEATURE
8-147 #domain translation elongation factor Tu homology #label
EF1
SUMMARY #length 428 #molecular-weight 47525 #checksum 4984

Query Match 54.4%; Score 62; DB 2; Length 428;
Best Local Similarity 57.1%; Pred. No. 1.18e+00;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 366 TCNIVEENPQFIKT 379
Qy 182 TSDTVNQNPQFIKA 195
I::I:|||||:

RESULT 3
ENTRY #type complete
TITLE translation elongation factor aEF-1 alpha chain -
Thermococcus celer
ORGANISM #formal_name Thermococcus celer
DATE 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
28-Aug-1998

ACCESSIONS S10248
REFERENCE Auer, J.; Spicker, G.; Boeck, A.
#authors Nucleic Acids Res. (1990) 18:3989
#journal Nucleotide sequence of the gene for elongation factor
#title EF-1-alpha from the extreme thermophilic archaeobacterium
Thermococcus celer.

#cross-references MUID:90326525
#accession S10248
#molecule_type DNA
#residues 1-428 #label AUE
#cross-references EMBL:X52383; NID:g48132; PID:g48133

GENETICS
#gene EF-1a
CLASSIFICATION #superfamily translation elongation factor Tu; translation
elongation factor Tu homology
KEYWORDS GTP binding; P-loop; protein biosynthesis
FEATURE
8-147 #domain translation elongation factor Tu homology #label
EF1\
14-21 #region nucleotide-binding motif A (P-loop)\
144-147 #region GTP-binding NKXD motif
SUMMARY #length 428 #molecular-weight 47505 #checksum 3819

Query Match 54.4%; Score 62; DB 2; Length 428;
Best Local Similarity 57.1%; Pred. No. 1.18e+00;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 366 TCNIVEENPQFIKT 379
Qy 182 TSDTVNQNPQFIKA 195
I::I:|||||:

RESULT 4
ENTRY #type complete
TITLE translation elongation factor aEF-1 alpha chain - Pyrococcus
woesei
ORGANISM #formal_name Pyrococcus woesei
```

```
DATE 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
28-Aug-1998
ACCESSIONS S19000
REFERENCE Creti, R.; Citarella, F.; Tiboni, O.; Sanangelantoni, A.;
Palm, P.; Cammarano, P.;
#authors J. Mol. Evol. (1991) 33:332-342
#journal Nucleotide sequence of a DNA region comprising the gene for
#title elongation factor 1-alpha (EF-1-alpha) from the
ultrathermophilic archaeote Pyrococcus woesei: phylogenetic
implications.

#cross-references MUID:92130258
#accession S19000
#molecule_type DNA
#residues 1-430 #label CRE
#cross-references EMBL:X59857; NID:g45946; PID:g45947
CLASSIFICATION #superfamily translation elongation factor Tu; translation
elongation factor Tu homology
KEYWORDS GTP binding; P-loop; protein biosynthesis
FEATURE
10-149 #domain translation elongation factor Tu homology #label
EF1\
16-23 #region nucleotide-binding motif A (P-loop)\
146-149 #region GTP-binding NKXD motif
SUMMARY #length 430 #molecular-weight 47923 #checksum 4204

Query Match 54.4%; Score 62; DB 2; Length 430;
Best Local Similarity 57.1%; Pred. No. 1.18e+00;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 368 TGNIVEENPQFIKT 381
Qy 182 TSDTVNQNPQFIKA 195
I::I:|||||:

RESULT 5
ENTRY #type complete
TITLE hypothetical protein YER128w - yeast (Saccharomyces
cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change
05-Dec-1997

ACCESSIONS S43222; S50631
REFERENCE Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.;
Kemp, C.; Wei, Y.; Taylor, P.; Nakahara, K.; Roberts, D.;
Davis, R.W.
#submission submitted to the EMBL Data Library, February 1993
#accession S43222
#molecule_type DNA
#residues 1-203 #label MUL
#cross-references GB:U18916; EMBL:L11119; NID:g1384128; PID:g603367
REFERENCE S50631
#authors Dietrich, F.S.
#submission submitted to the EMBL Data Library, December 1994
#description The sequence of S. cerevisiae cosmid 9781, 8198, 9115, 9981,
and lambda clones 3955 and 6052.
#accession S50631
#molecule_type DNA
#residues 1-203 #label DIE
#cross-references EMBL:U18916; NID:g1384128; PID:g603367; MIPS:YER128w
GENETICS
#map_position 5R
SUMMARY #length 203 #molecular-weight 23488 #checksum 4017

Query Match 53.5%; Score 61; DB 2; Length 203;
Best Local Similarity 53.3%; Pred. No. 1.80e+00;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Db 38 LYTCEIHLDQNPQFV 52
Qy 180 LYTSDTVNQ-NPQFI 193
I::I:|||||:
```



```

RESULT      6
ENTRY       T00697      #type complete
TITLE       probable glycopeptide N-tetradecanoyltransferase (EC
              2.3.1.97) - Arabidopsis thaliana
ALTERNATE_NAMES
ORGANISM     #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
DATE        01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
              01-Feb-1999
ACCESSIONS  T00697
REFERENCE    214198
#authors     Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby,
              M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage,
              A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
#submission submitted to the EMBL Data Library, June 1998
#description Arabidopsis thaliana chromosome II BAC F6E13 genomic
              sequence.
#accession  T00697
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    1-389 ##label ROU
#cross-references EMBL:AC004005; NID:g3212846; PID:g3212870
GENETICS
#map_position II
#introns     100/3
#note        F6E13.30
KEYWORDS     acyltransferase
SUMMARY      #length 389 #molecular-weight 44450 #checksum 5324
              Query Match 52.6%; Score 60; DB 2; Length 389;
              Best Local Similarity 60.0%; Pred. No. 2.73e+00;
              Matches 9; Conservative 3; Mismatches 2; Indels 1; Gaps 1;
Db          343 FYASD-VMQNESFLK 356
              :||| |||| |||
Qy          180 LYTSDTMQNPQFIK 194
              ||| :||| |||

RESULT      7
ENTRY       A49810      #type complete
TITLE       hypothetical protein (gene 7 5' region) - rotavirus B
              #formal_name rotavirus B
ORGANISM     01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change
              09-Sep-1997
DATE        A49810
REFERENCE    A49810
#authors     Eiden, J.J.
#journal     Virology (1994) 199:212-218
#title       Expression and sequence analysis of gene 7 of the IDIR agent
              (group B rotavirus): similarity with NS53 of group A
              rotavirus.
#cross-references MUID:94160571
#accession  A49810
#status      preliminary
#molecule_type genomic RNA
#residues    1-115 ##label EID
#cross-references GB:U01164; NID:g470378; PID:g623030
SUMMARY      #length 115 #molecular-weight 12845 #checksum 4799
              Query Match 51.8%; Score 59; DB 2; Length 115;
              Best Local Similarity 53.3%; Pred. No. 4.13e+00;
              Matches 8; Conservative 2; Mismatches 4; Indels 1; Gaps 1;
Db          92 LYV-QPVMQSNPFIF 105
              || :||| |||
Qy          180 LYTSDTMQNPQFIK 194
              ||| :||| |||

RESULT      8
ENTRY       T02140      #type complete
TITLE       hypothetical protein F8K4.15 - Arabidopsis thaliana
              #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
ALTERNATE_NAMES
ORGANISM     #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
DATE        05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
              05-Mar-1999
ACCESSIONS  T02140
REFERENCE    214574
#authors     Vysotskaia, V.S.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu,
              G.; Oji, O.; Liu, S.; Li, J.; Araujo, R.; Au, M.; Brendel,
              V.; Buehler, E.; Conway, A.B.; Conway, A.R.; Dewar, K.;
              Feng, J.; Kim, C.; Kurtz, D.; Li, Y.; Palm, C.J.; Shinn,
              P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N.A.;
              Theologis, A.
#submission submitted to the EMBL Data Library, August 1998
#description Arabidopsis thaliana chromosome I BAC F8K4 sequence.
#accession  T02140
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    1-457 ##label VYS
#cross-references EMBL:AC004392; NID:g3282170; PID:g3367526
GENETICS
#map_position 1
#note        F8K4.15
SUMMARY      #length 457 #molecular-weight 51477 #checksum 4674
              Query Match 51.8%; Score 59; DB 2; Length 457;
              Best Local Similarity 46.2%; Pred. No. 4.13e+00;
              Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
Db          358 YTAETVKKKTEFI 370
              |||:|||| :|||
Qy          181 YTSDTVMQNPQFI 193
              |||:|||| :|||

RESULT      9
ENTRY       A60179      #type fragment
TITLE       3',5'-cyclic-nucleotide phosphodiesterase (EC 3.1.4.17),
              cyclic GMP-stimulated, brain - bovine (fragment)
ORGANISM     #formal_name Bos primigenius taurus #common_name cattle
              03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change
              17-Mar-1999
DATE        A60179
REFERENCE    A60179
#authors     Tanaka, T.; Hockman, S.; Moos Jr., M.; Taira, M.; Meaccci, E.;
              Murashima, S.; Manganelli, V.C.
#journal     Second Messengers Phosphoproteins (1991) 13:97-98
#title       Comparison of putative cGMP-binding regions in bovine brain
              and cardiac cGMP-stimulated phosphodiesterases.
#cross-references MUID:92065414
#accession  A60179
#status      not compared with conceptual translation
#molecule_type mRNA
#residues    1-611 ##label TAN
#note        part of this sequence was confirmed by protein
              sequencing
KEYWORDS     cGMP binding; homodimer; phosphoric diester hydrolase
FEATURE      81-579
SUMMARY      #domain cyclic-nucleotide phosphodiesterase similarity
              #label CNP
              #length 611 #checksum 9356
              Query Match 51.8%; Score 59; DB 2; Length 611;
              Best Local Similarity 41.2%; Pred. No. 4.13e+00;
              Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;
Db          388 LYSSEGSVMRRHFAQA 404
              |||:|||| :|||
Qy          180 LYTSD-TVMQNPQFIKA 195
              |||:|||| :|||

RESULT      10
ENTRY       JC4397      #type complete
TITLE       peroxinectin precursor - signal crayfish
              cell adhesion protein
ALTERNATE_NAMES
CONTAINS     peroxidase (EC 1.11.1.7)
```

ORGANISM	#formal_name Pacifastacus leniusculus #common_name signal crayfish
DATE	20-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 17-Mar-1999
ACCESSIONS	JC4397: PC4113
REFERENCE	JC4397
#authors	Johansson, M.W.; Lind, M.I.; Holmblad, T.; Thoenqvist, P.O.; Soederhaell, K.
#journal	Biochem. Biophys. Res. Commun. (1995) 216:1079-1087
#title	Peroxinectin, a novel cell adhesion protein from crayfish blood.
#cross-references	MUID:96074593
#accession	JC4397
#molecule_type	mRNA
##residues	1-818 ##label JOH
##cross-references	EMBL:X91409: NID:gl150531; PID:e199227; PID:gl150532
##experimental_source	Blood cells
#accession	PC4113
#molecule_type	protein
##residues	185-188; 660:662-667 ##label JO2
COMMENT	This protein acts as both cell adhesion ligand and peroxidase, and it belongs to a family of proteins that includes myeloperoxidase. It also triggers an intracellular signaling pathway involving protein kinase C activation and protein tyrosine phosphorylation, leading to cell spreading and degranulation, stimulates cellular reactions.
CLASSIFICATION	superfamily peroxinectin; myeloperoxidase homology
KEYWORDS	blood; cell adhesion; chromoprotein; glycoprotein; heme; iron; oxidoreductase
FEATURE	
1-18	#domain signal sequence #status predicted #label SIG\
19-818	#product peroxinectin #status predicted #label MAT\
143-808	#domain myeloperoxidase homology #label MPX\
739-741	#region cell attachment #status predicted\
235-254, 346-362,	
347-380, 453-462,	#disulfide_bonds #status predicted\
672-728, 772-798,	#binding_site heme iron (His) (axial ligands) #status predicted\
328, 573	
349	#binding_site carbohydrate (Asn) (covalent) #status predicted\
469	#active_site Arg #status predicted
SUMMARY	#length 818 #molecular_weight 89315 #checksum 1903
Query Match	51.8%; Score 59; DB 2; Length 818;
Best Local Similarity	64.3%; Pred No. 4.13e+00;
Matches	9; Conservative 3; Mismatches 1; Indels 1; Gaps 1;
Db	314 LFTS-SYMQWAQFI 326
Qy	180 LYTSDTVMQNPQFI 193
RESULT	11
ENTRY	A40981 #type complete
TITLE	3',5'-cyclic-nucleotide phosphodiesterase (BC 3.1.4.17), cyclic GMP-stimulated - bovine
ALTERNATE_NAMES	cGMP-dependent phosphodiesterase
ORGANISM	#formal_name Bos primigenius taurus #common_name cattle
DATE	03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change 10-Sep-1997
ACCESSIONS	A40981; B36112; A36112; B26650; C26650
REFERENCE	A40981
#authors	Sonnenburg, W.K.; Mullaney, P.J.; Beavo, J.A.
#journal	J. Biol. Chem. (1991) 266:17653-17661
#title	Molecular cloning of a cyclic GMP-stimulated cyclic nucleotide phosphodiesterase cDNA. Identification and distribution of isozyme variants.
#cross-references	MUID:91373395
#accession	A40981
#molecule_type	mRNA
##residues	1-921 ##label SON
##cross-references	GB:M73512: NID:gl62829; PID:gl62830

```
195,411,590,692,
898      #binding_site carbohydrate (asn) (covalent) #status
SUMMARY      predicted
#length 928 #molecular-weight 104663 #checksum 7548
Query Match      51.8%; Score 59; DB 2; Length 928;
Best Local Similarity 41.2%; Pred. No. 4.13e+00;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 706 LYSSEGSVNERHHFAQA 722
|||: |||: |||
Qy 180 LYTSD-TVMQNPFQFIKA 195

RESULT 13
ENTRY      #type complete
TITLE      ATP-dependent nuclease - Helicobacter pylori (strain 26695)
ORGANISM    #formal_name Helicobacter pylori
DATE        09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
10-Oct-1997
ACCESSIONS C64554
REFERENCE   A64520
#authors    Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.;
McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
Fuji, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
C.M.; Venter, J.C.
#journal    Nature (1997) 388:539-547
#title      The complete genome sequence of the gastric pathogen
Helicobacter pylori.
#cross-references MUID:97394467
#accession  C64554
#status     preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues   1-430 #label TOM
#cross-references GB:AE000546; GB:AE000511; NID:G2313363; PID:G2313366;
TIGR:RP0275
SUMMARY      #length 430 #molecular-weight 49684 #checksum 1022
Query Match      50.9%; Score 58; DB 2; Length 430;
Best Local Similarity 46.2%; Pred. No. 6.22e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 223 LYKKNPIVONAQF 235
|||: |||: |||
Qy 180 LYTSDTVMNQPF 192

RESULT 14
ENTRY      #type complete
TITLE      hypothetical protein jhp0260 - Helicobacter pylori (strain
J99)
ORGANISM    #formal_name Helicobacter pylori
DATE        12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
12-Feb-1999
ACCESSIONS C71954
REFERENCE   A71800
#authors    Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal    Nature (1999) 397:176-180
#title      Genomic sequence comparison of two unrelated isolates of the
```

```
human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#accession  C71954
#status     preliminary
#molecule_type DNA
#residues   1-431 #label ARN
#cross-references GB:AE001463; GB:AE001439; NID:G4154775; PID:G4154781
#experimental_source strain J99
GENETICS
#gene       jhp0260
SUMMARY      #length 431 #molecular-weight 49895 #checksum 5304
Query Match      50.9%; Score 58; DB 2; Length 431;
Best Local Similarity 46.2%; Pred. No. 6.22e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 224 LYKKNPIVONAQF 236
|||: |||: |||
Qy 180 LYTSDTVMNQPF 192

RESULT 15
ENTRY      #type complete
TITLE      probable membrane protein YDR141c - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES hypothetical protein YD9302.17c
ORGANISM    #formal_name Saccharomyces cerevisiae
DATE        05-May-1995 #sequence_revision 21-Jul-1995 #text_change
12-Dec-1997
ACCESSIONS S51869
REFERENCE   S51853
#authors    Oliver, K.; Harris, D.
#submission submitted to the EMBL Data Library, February 1995
#accession  S51869
#molecule_type DNA
#residues   1-1698 #label OLI
#cross-references EMBL:Z48179; NID:G665657; PID:G665674; MIPS:YDR141c
GENETICS
#map_position 4R
KEYWORDS     transmembrane protein
FEATURE
496-512      #domain transmembrane #status predicted #label TM1\
943-959      #domain transmembrane #status predicted #label TM2\
1285-1301    #domain transmembrane #status predicted #label TM3
SUMMARY      #length 1698 #molecular-weight 194685 #checksum 8682
Query Match      50.9%; Score 58; DB 2; Length 1698;
Best Local Similarity 43.8%; Pred. No. 6.22e+00;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 436 LYSSEKFKQNEQIMKT 451
|||: |||: |||
Qy 180 LYTSDTVMNQPFQFIKA 195

Search completed: Thu Jul 8 18:28:14 1999
Job time : 11 secs.
```

THIS PAGE BLANK (USPTO)


```
Best Local Similarity 37.5%; Pred. No. 9.25e-02;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Db 189 LYSDEAAVQAPTFLRA 204
    ||::: ||| |||
Qy 180 LYTSDTVMNQPFQFIKA 195

RESULT 2
ID EF1A_THECE STANDARD; PRT; 428 AA.
AC P17197;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) (ELONGATION FACTOR TU) (EF-TU).
GN TUF.
OS THERMOCOCCUS CELER.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; THERMOCOCCUS.
RN [1]
RX MEDLINE; 90326525.
RA AUER J., SPICKER G., BOECK A.;
RT "Nucleotide sequence of the gene for elongation factor EF-1 alpha
RL NUCLEIC ACIDS RES. 18:3989-3989(1990).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB009517; D1028591; -
DR PROSITE; PS00301; EFACOR_GTP; 1.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 89 93 GTP (BY SIMILARITY).
FT NP_BIND 144 147 GTP (BY SIMILARITY).
SQ SEQUENCE 428 AA; 47505 MW; 97BB3B2D CRC32;

Query Match 54.4%; Score 62; DB 1; Length 428;
Best Local Similarity 57.1%; Pred. No. 3.84e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 366 TGNVEENPQFIKT 379
    ||::: ||| |||
Qy 182 TSDTVMNQPFQFIKA 195

RESULT 4
ID EF1A_PYRMO STANDARD; PRT; 430 AA.
AC P26751;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) (ELONGATION FACTOR TU) (EF-TU).
GN TUF.
OS PYROCOCUS WOESSEI.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCUS.
RN [1]
RX MEDLINE; 92130258.
RA CRETJ R., CITARELLA F., TIBONI O., SANANGELANTONI A.M., PALM P.,
RA CAMMARANO P.;
RT "Nucleotide sequence of a DNA region comprising the gene for
RT elongation factor 1 alpha (EF-1 alpha) from the ultrathermophilic
RT archaeote Pyrococcus woesei: phylogenetic implications.";
RL J. MOL. EVOL. 33:332-342(1991).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52383; G48133; -
DR PIR; S10248;
DR PROSITE; PS00301; EFACOR_GTP; 1.
DR PFAM; PF00009; GTP_EFTU; 1.
DR HSP; P07157; 1AIP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 89 93 GTP (BY SIMILARITY).
FT NP_BIND 144 147 GTP (BY SIMILARITY).
SQ SEQUENCE 428 AA; 47505 MW; 97BB3B2D CRC32;

Query Match 54.4%; Score 62; DB 1; Length 428;
Best Local Similarity 57.1%; Pred. No. 3.84e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 366 TGNVEENPQFIKT 379
    ||::: ||| |||
Qy 182 TSDTVMNQPFQFIKA 195

RESULT 3
ID EF1A_PYRHO STANDARD; PRT; 428 AA.
AC OS9153;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) (ELONGATION FACTOR TU) (EF-TU).
GN TUF OR PHC0033.
OS PYROCOCUS HORIKOSHII.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCUS.
RN [1]
RX MEDLINE; 90326525.
RA AUER J., SPICKER G., BOECK A.;
RT "Nucleotide sequence of the gene for elongation factor EF-1 alpha
RL NUCLEIC ACIDS RES. 18:3989-3989(1990).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52383; G48133; -
DR PIR; S10248;
DR PROSITE; PS00301; EFACOR_GTP; 1.
DR PFAM; PF00009; GTP_EFTU; 1.
DR HSP; P07157; 1AIP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 89 93 GTP (BY SIMILARITY).
FT NP_BIND 144 147 GTP (BY SIMILARITY).
SQ SEQUENCE 428 AA; 47505 MW; 97BB3B2D CRC32;

Query Match 54.4%; Score 62; DB 1; Length 428;
Best Local Similarity 57.1%; Pred. No. 3.84e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 366 TGNVEENPQFIKT 379
    ||::: ||| |||
Qy 182 TSDTVMNQPFQFIKA 195
```

```
CC or send an email to license@isb-sib.ch).
DE -----
GN EMBL: X59857; G45947; -
OS PIR: S19000; S19000.
OC PFAM: PF00301; EFATOR_GTP; 1.
OC PFAM: PF00009; GTP_EFTU; 1.
RN HSSP: P07157; LAIP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 16 23 GTP (BY SIMILARITY).
FT NP_BIND 91 95 GTP (BY SIMILARITY).
FT NP_BIND 146 149 GTP (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47923 MW; 2F1EDDD1 CRC32;

Query Match
Best Local Similarity 54.1%; Score 62; DB 1; Length 430;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 368 TGNVEENPOFIKT 381
QY 182 TSDTVNQNPQFIKA 195
::: | : |||||:

RESULT 5
ID YE88_YEAST STANDARD; PRT; 203 AA.
AC P40080;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 23.5 KD PROTEIN IN RSP5-PAK1 INTERGENIC REGION.
GN YER128W OR SYGP-ORF44.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHUNG E., DUNCAN A., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U18916; G603367; -
DR PIR: S43222; S43222.
KW HYPOTHETICAL PROTEIN
SQ SEQUENCE 203 AA; 23488 MW; 9136E399 CRC32;

Query Match
Best Local Similarity 53.5%; Score 61; DB 1; Length 203;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Db 38 LYTCEHLDONPOFV 52
QY 180 LYTSDTVMQ-NPQFI 193
::: | : |||||:

RESULT 6
ID CN2A_BOVIN STANDARD; PRT; 921 AA.
AC P14099; Q28064;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
```

```
DE CGMP-DEPENDENT 3',5'-CYCLIC PHOSPHODIESTERASE (EC 3.1.4.17)
GN (CYCLIC GMP STIMULATED PHOSPHODIESTERASE) (CGS-PDE).
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINE; BOS.
RN [1]
RP SEQUENCE FROM N.A. (PDE2A1).
RX MEDLINE; 91373395.
RA SONNENBURG W.K., MULLANEY P.J., BEAVO J.A.;
RT "Molecular cloning of a cyclic GMP-stimulated cyclic nucleotide
RT phosphodiesterase cDNA. Identification and distribution of isozyme
RT variants.";
RL J. BIOL. CHEM. 266:17655-17661(1991).
RN [2]
RP SEQUENCE FROM N.A. (PDE2A3).
RC TISSUE=BRAIN;
RA JUILFS D.M., SONNENBURG W.K., SERAJI S., BEAVO J.A.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 592-921 FROM N.A.
RX TISSUE=HEART;
RX MEDLINE; 91104948.
RA LE TRONG H., BEIER N., SONNENBURG W.K., STROOP S.D., WALSH K.A.,
RA BEAVO J.A., CHARBONNEAU H.;
RT "Amino acid sequence of the cyclic GMP stimulated cyclic nucleotide
RT phosphodiesterase from bovine heart.";
RL BIOCHEMISTRY 29:10280-10288(1990).
RN [4]
RP SEQUENCE OF 613-694 AND 808-868.
RC TISSUE=HEART;
RX MEDLINE; 87092242.
RA CHARBONNEAU H., BEIER N., WALSH K.A., BEAVO J.A.;
RT "Identification of a conserved domain among cyclic nucleotide
RT phosphodiesterases from diverse species.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:9308-9312(1986).
CC -!- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O =
CC GUANOSINE 5'-PHOSPHATE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC -!- ALTERNATIVE PRODUCTS: AT LEAST THREE FORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING. THE SEQUENCE SHOWN HERE IS THAT OF ISOFORM
CC PDE2A1.
CC -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M73512; G162830; -
DR EMBL: L49503; G1184300; -.
DR PIR: A40981; A40981.
DR PFAM: PF00126; PDEASE_I; 1.
DR PFAM: PF00233; PDEase; 1.
KW HYDROLASE; CGMP; ALTERNATIVE SPLICING; MEMBRANE; ACETYLATION.
FT MOD_RES 1 1 ACETYLATION.
FT DOMAIN 613 871 CATALYTIC.
FT VARSPPLIC 1 25 MRQPAARSDRLFQAEPVPPGSDGA -> MGQACGHSILCR
FT SOQYPAARPAEPARQGVFLKPDPPPPPPPCADS (IN
FT PDE2A3).
FT CONFLICT 204 204 N -> D (IN REF. 2).
FT CONFLICT 633 633 P -> L (IN REF. 4).
SQ SEQUENCE 921 AA; 103227 MW; 000F1172 CRC32;

Query Match
Best Local Similarity 51.8%; Score 59; DB 1; Length 921;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;
```

```
Db 698 LYSSEGSVMRHHFAQA 714
||:|: ||: |
Qy 180 LYTSD-TVMQNPOFIKA 195

RESULT 7
ID CNZA_RAT STANDARD; PRT; 928 AA.
AC Q01062;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 36, LAST ANNOTATION UPDATE)
DE CGMP-DEPENDENT 3',5'-CYCLIC PHOSPHODIESTERASE (EC 3.1.4.17)
DE (CYCLIC GMP STIMULATED PHOSPHODIESTERASE) (CGS-PDE).
GN PDE2A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN: SPRAGUE-DAWLEY; TISSUE: BRAIN;
RX MEDLINE: 95110334.
RA YANG Q., PASKIND M., BOLGER G., THOMPSON W.J., REPASKE D.R.,
RA CUTLER L.S., EPSTEIN P.M.;
RT "A novel cyclic GMP stimulated phosphodiesterase from rat brain.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 205:1850-1858(1994).
RN [2]
RP SEQUENCE OF 643-759 FROM N.A.
RX MEDLINE: 92406782.
RT REPASKE D.R., SWINNEN J.V., JIN S.L.C., VAN WYK J.J., CONTI M.;
RT "A polymerase chain reaction strategy to identify and clone cyclic
nucleotide phosphodiesterase cDNAs. Molecular cloning of the cDNA
encoding the 63-kDa calmodulin-dependent phosphodiesterase.";
RL J. BIOL. CHEM. 267:18683-18688(1992).
CC -1- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O =
GUANOSINE 5'-PHOSPHATE.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST THREE FORMS ARE PRODUCED BY
ALTERNATIVE SPLICING. THE SEQUENCE SHOWN HERE IS THAT OF ISOFORM
PDE2A2.
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: U21101; G706930; -
DR EMBL: M94540; G203446; -
DR PROSITE: PS00126; PDEASE_I; 1.
DR PFAM: PF00233; PDEase; 1.
KW HYDROLASE; CGMP; ALTERNATIVE SPLICING; MEMBRANE.
SQ SEQUENCE 928 AA; 104663 MW; 6D2390BD CRC32;
Query Match 51.8%; Score 59; DB 1; Length 928;
Best Local Similarity 41.2%; Pred. No. 1.53e+00;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 706 LYSSEGSVMRHHFAQA 722
||:|: ||: |
Qy 180 LYTSD-TVMQNPOFIKA 195

RESULT 8
ID CNZA_HUMAN STANDARD; PRT; 941 AA.
AC O00408;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 36, LAST ANNOTATION UPDATE)
DE CGMP-DEPENDENT 3',5'-CYCLIC PHOSPHODIESTERASE (EC 3.1.4.17)
DE (CYCLIC GMP STIMULATED PHOSPHODIESTERASE) (CGS-PDE).
GN PDE2A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN: SPRAGUE-DAWLEY; TISSUE: BRAIN;
RX MEDLINE: 95110334.
RA YANG Q., PASKIND M., BOLGER G., THOMPSON W.J., REPASKE D.R.,
RA CUTLER L.S., EPSTEIN P.M.;
RT "A novel cyclic GMP stimulated phosphodiesterase from rat brain.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 205:1850-1858(1994).
RN [2]
RP SEQUENCE OF 643-759 FROM N.A.
RX MEDLINE: 92406782.
RT REPASKE D.R., SWINNEN J.V., JIN S.L.C., VAN WYK J.J., CONTI M.;
RT "A polymerase chain reaction strategy to identify and clone cyclic
nucleotide phosphodiesterase cDNAs. Molecular cloning of the cDNA
encoding the 63-kDa calmodulin-dependent phosphodiesterase.";
RL J. BIOL. CHEM. 267:18683-18688(1992).
CC -1- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O =
GUANOSINE 5'-PHOSPHATE.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST THREE FORMS ARE PRODUCED BY
ALTERNATIVE SPLICING. THE SEQUENCE SHOWN HERE IS THAT OF ISOFORM
PDE2A2.
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: U21101; G706930; -
DR EMBL: M94540; G203446; -
DR PROSITE: PS00126; PDEASE_I; 1.
DR PFAM: PF00233; PDEase; 1.
KW HYDROLASE; CGMP; ALTERNATIVE SPLICING; MEMBRANE.
FT CONFLICT 646 646 W -> R (IN REF. 2).
FT CONFLICT 758 758 L -> M (IN REF. 2).
SQ SEQUENCE 928 AA; 104663 MW; 6D2390BD CRC32;
Query Match 51.8%; Score 59; DB 1; Length 928;
Best Local Similarity 41.2%; Pred. No. 1.53e+00;
Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 706 LYSSEGSVMRHHFAQA 722
||:|: ||: |
Qy 180 LYTSD-TVMQNPOFIKA 195

RESULT 9
ID KINH_STRPU STANDARD; PRT; 1031 AA.
AC P35978;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE KINESIN HEAVY CHAIN.
OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC STRONGYLOCENTROTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91225077.
RA WRIGHT B.D., HENSON J.H., WEDAMAN K.P., WILLY P.J., MORAND J.N.,
RA SCHOLEY J.M.;
RT "Subcellular localization and sequence of sea urchin kinesin heavy
chain: evidence for its association with membranes in the mitotic
apparatus and interphase cytoplasm.";
RL J. CELL BIOL. 113:817-833(1991).
CC -1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
```



```
CC -1- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC CHAINS.
CC -1- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC VESICLES AND MEMBRANOUS ORGANELLES.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X56844; GI0270; -.
CC PIR: A38713; A38713.
CC PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC PFAM: PF00225; kinesin; 1.
CC HSSP: P56536; 2KIN.
CC MOTOR PROTEIN; MICROTUBULES; ATP-BINDING; COILED COIL.
KW DOMAIN 1 384 MECHANOCHEMICAL (MOTOR) (BY SIMILARITY).
FT DOMAIN 393 857 COILED COIL.
FT DOMAIN 858 1031 GLOBULAR.
FT NP_BIND 84 91 ATP (BY SIMILARITY).
SQ SEQUENCE 1031 AA; 117523 MW; B0DC4D39 CRC32;

Query Match 50.0%; Score 57; DB 1; Length 1031;
Best Local Similarity 46.2%; Pred. No. 3.72e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 92 FTMEGVGNPQYM 104
QY 181 YTSDTVMQNPFQI 193

RESULT 10
ID ALFC_SPTOL STANDARD; PRT; 394 AA.
AC P16096;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DE FRUCTOSE-BISPHOSPHATE ALDOLASE, CHLOROPLAST PRECURSOR (EC 4.1.2.13).
OS SPINACIA OLERACEA (SPINACH).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMERYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC CARYOPHYLLIDAE; CARYOPHYLLALES; CHENOPODIACEAE; SPINACIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93144707.
RA PELZER-REITH B., PENDER A., SCHNARENBERGER C.;
RT "Plant aldolase: cDNA and deduced amino-acid sequences of the
RT chloroplast and cytosol enzyme from spinach.";
RL PLANT MOL. BIOL. 21:331-340(1993).
RN [2]
RP SEQUENCE OF 47-64.
RX MEDLINE: 8411487.
RA LEBHERZ H.G., LEADBETTER M.M., BRADSHAW R.A.;
RT "Isolation and characterization of the cytosolic and chloroplast
RT forms of spinach leaf fructose diphosphate aldolase.";
RL J. BIOL. CHEM. 259:1011-1017(1984).
RN [3]
RP IDENTIFICATION OF PROBABLE FRAMESHIFT.
RA BAIRCH A.;
RL UNPUBLISHED OBSERVATIONS (NOV-1995).
CC -1- CATALYTIC ACTIVITY: D-FRUCTOSE 1,6-BISPHOSPHATE = GLYCERONE-
CC PHOSPHATE + D-GLYCERALDEHYDE 3-PHOSPHATE.

CC -1- PATHWAY: SIXTH STEP IN GLYCOLYSIS.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- SIMILARITY: BELONGS TO CLASS I FRUCTOSE-BISPHOSPHATE ALDOLASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X66814; G22633; -.
CC PIR: S22092; ADSAP.
CC PIR: A21815; A21815.
CC PROSITE: PS00158; ALDOLASE_CLASS_I; 1.
CC PFAM: PF00274; glycolytic_enz; 1.
CC HSSP: P14223; 1A5C.
CC LYASE; SCHIFF BASE; GLYCOLYSIS; CHLOROPLAST; TRANSIT PEPTIDE.
KW TRANSIT 1 46 CHLOROPLAST.
FT CHAIN 47 394 FRUCTOSE-BISPHOSPHATE ALDOLASE.
FT BINDING 93 93 C-1-PHOSPHATE GROUP OF THE SUBSTRATE
FT (BY SIMILARITY).
FT BINDING 183 183 C-1-PHOSPHATE GROUP OF THE SUBSTRATE
FT (BY SIMILARITY).
FT BINDING 265 265 SCHIFF-BASE WITH DIHYDROXYACETONE-P
FT (BY SIMILARITY).
FT ACT_SITE 394 394 ESSENTIAL FOR ENHANCED ACTIVITY OF THE
FT ENZYME TOWARD FRUCTOSE 1,6-BISPHOSPHATE
FT AS COMPARED WITH FRUCTOSE 1-PHOSPHATE
FT (BY SIMILARITY).
FT CONFLICT 260 263 EGSS -> RDP (IN REF. 1).
SQ SEQUENCE 394 AA; 42468 MW; 82562E7B CRC32;

Query Match 49.1%; Score 56; DB 1; Length 394;
Best Local Similarity 37.5%; Pred. No. 5.76e+00;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Db 6 LKKTSPVLDNPEFLKG 21
QY 180 LYTSDTVMQNPFQFI 195

RESULT 11
ID SLAL_YEAST STANDARD; PRT; 1244 AA.
AC P32790;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CYTOSKELETON ASSEMBLY CONTROL PROTEIN SLAL.
GN SLAL OR YBL007C OR YBL0321.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BDY 228;
RX MEDLINE: 93328765.
RA HOLTZMAN D.A., YANG S., DRUBIN D.G.;
RT "Synthetic-lethal interactions identify two novel genes, SLAL and
RT SLA2, that control membrane cytoskeleton assembly in Saccharomyces
RT cerevisiae.";
RL J. CELL BIOL. 122:635-644(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C;
RX MEDLINE: 93070613.
RA DELAVEAU T., JACO C., PEREA J.;
RT "Sequence of a 12.7 kb segment of yeast chromosome II identifies a
RT PDR-like gene and several new open reading frames.";
RL YEAST 8:761-768(1992).
CC -1- FUNCTION: ESSENTIAL FOR THE PROPER FORMATION OF THE CORTICAL
```

```
CC ACTIN CYTOSKELETON. INVOLVED IN CONTROLLING THE SIZE OF CORTICAL
CC PATCHES PERHAPS BY REGULATING THE NUCLEATION OF FILAMENTS AT THE
CC CORTEX.
CC -!- SIMILARITY: SOME, TO SEA URCHIN BINDIN.
CC -!- SIMILARITY: CONTAINS 3 SH3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z22810; G311411; -.
CC DR EMBL: Z35768; G535990; -.
CC DR EMBL: S47695; E65054; -.
CC DR PIR: S25327; S25327.
CC DR PIR: A40673; A40673.
CC DR PIR: S36355; S36355.
CC DR SGD: L0001912; SLA1.
CC DR PROSITE: PS50002; SH3; 3.
CC DR PFAM: PF00018; SH3; 3.
CC DR HSSP: Q03526; LAWJ.
CC KW CYTOSKELETON; ACTIN-BINDING; SH3 DOMAIN; REPEAT.
CC FT DOMAIN 8 69
CC FT DOMAIN 70 132
CC FT DOMAIN 353 415
CC FT DOMAIN 868 1205
CC FT REPEAT 868 874 1.
CC FT REPEAT 877 883 2.
CC FT REPEAT 887 893 3.
CC FT REPEAT 923 929 4.
CC FT REPEAT 945 951 5.
CC FT REPEAT 1003 1009 6.
CC FT REPEAT 1020 1026 7.
CC FT REPEAT 1031 1037 8.
CC FT REPEAT 1048 1054 9.
CC FT REPEAT 1065 1071 10.
CC FT REPEAT 1084 1090 11.
CC FT REPEAT 1129 1135 12.
CC FT REPEAT 1155 1161 13.
CC FT REPEAT 1170 1176 14.
CC FT REPEAT 1185 1191 15.
CC FT REPEAT 1200 1206 16.
CC SQ SEQUENCE 1244 AA; 135848 MW; 6A281130 CRC32;
Query Match 49.1%; Score 56; DB 1; Length 1244;
Best Local Similarity 33.3%; Pred. No. 5.76e+00;
Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
Db 1183 FNTDGMQPMQMMNT 1197
: :| :||| :|
Qy 181 YTSDTVMQNPFKA 195
RESULT 12
ID LECALOR STANDARD; PRT; 109 AA.
AC P49329;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE MANNOSSE-SPECIFIC LECTIN (AGGLUTININ).
OS ALOE ARBORESCENS (KIDACHI ALOE).
OC EUKARYOTA; VIRIDIPALTA; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; LILIALES;
OC ALOACEAE; ALOE.
RN [1]
RP SEQUENCE.
RC STRAIN-CV. NATALENSIS BERGER; TISSUE=LEAF;
RX MEDLINE: 95398630.
RA KOIKE T., TITANI K., SUZUKI M., BEPPU H., KUZUYA H., MARUTA K.,
```

```
RA SHIMPO K., FUJITA K.;
RT "The complete amino acid sequence of a mannose-binding lectin from
RT 'Kidachi Aloe' (Aloe arborescens Miller var. natalensis Berger).";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 214:163-170(1995).
RN [2]
RP SEQUENCE OF 1-24 AND 83-106.
RC STRAIN-CV. NATALENSIS BERGER; TISSUE=LEAF;
RX MEDLINE: 96362159.
RA KOIKE T., BEPPU H., KUZUYA H., MARUTA K., SHIMPO K., SUZUKI M.,
RA TITANI K., FUJITA K.;
RT "A 35 kDa mannose-binding lectin with hemagglutinating and mitogenic
RT activities from 'Kidachi Aloe' (Aloe arborescens Miller var.
RT natalensis Berger).";
RL J. BIOCHEM. 118:1205-1210(1995).
CC -!- FUNCTION: MANNOSSE-SPECIFIC LECTIN. SHOWS AGGLUTINATING ACTIVITY
CC TOWARD RABBIT ERYTHROCYTES AND MITOGENIC ACTIVITY TOWARDS MOUSE
CC LYMPHOCYTES.
CC -!- SUBUNIT: HOMOTRIMER OR HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: TO G. NIVALIS MANNOSSE-SPECIFIC LECTIN.
DR HSSP: P30617; 1NIV.
KW LECTIN; REPEAT: MANNOSSE-BINDING.
FT CHAIN 1 78 MANNOSSE-SPECIFIC LECTIN HEAVY CHAIN.
FT PROPEP 79 82
FT CHAIN 83 109 MANNOSSE-SPECIFIC LECTIN LIGHT CHAIN.
FT DISULFID 29 52 V -> I.
FT VARIANT 63 63 N -> F.
FT VARIANT 76 76 N -> D.
FT VARIANT 94 94 A -> S.
FT VARIANT 104 104 2.5 X APPROXIMATE REPEATS.
FT DOMAIN 25 106 1.
FT REPEAT 25 47 1.
FT REPEAT 56 80 2.
FT REPEAT 89 106 3 (INCOMPLETE).
SQ SEQUENCE 109 AA; 11941 MW; 1111FC4A CRC32;
Query Match 48.2%; Score 55; DB 1; Length 109;
Best Local Similarity 42.9%; Pred. No. 8.86e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 1; Gaps 1;
Db 4 LYSSE-VLHENQYI 16
||:| :| :| :|
Qy 180 LYSDTVMQNPOFI 193
RESULT 13
ID HBBA_BOSJA STANDARD; PRT; 145 AA.
AC P04346;
DT 20-MAR-1987 (REL. 04, CREATED)
DT 20-MAR-1987 (REL. 04, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HEMOGLOBIN BETA-A CHAIN.
OS BOS JAVANICUS (WILD BANTENG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVINAE; BOS.
RN [1]
RP SEQUENCE.
RC STRAIN=HEREFORD;
RX MEDLINE: 84023669.
RA NAMIKAWA T., TAKENAKA O., TAKAHASHI K.;
RT "Hemoglobin Bali (bovine): beta A 18(B1)lys leads to His: one of the
RT 'missing links' between beta A and beta B of domestic cattle exists
RT in the Bali cattle (Bovinae, Bos banteng).";
RL BIOCHEM. GENET. 21:787-796(1983).
DR PIR: A02388; HBBOBB.
DR PROSITE: PS01033; GLOBIN; 1.
DR PFAM: PF00042; globin; 1.
DR HSSP: P02070; 1HDA.
DR HEME: OXYGEN TRANSPORT; RESPIRATORY PROTEIN; ERYTHROCYTE.
KW METAL 52 IRON (HEME DISTAL LIGAND).
FT METAL 91 91 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 145 AA; 15964 MW; CDE149E4 CRC32;
```


THIS PAGE BLANK (USPTO)

WQRELH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:27:24 1999; MasPar time 5.61 Seconds
155.739 Million cell updates/sec
Tubular output not generated.

Title: >US-09-041-236-2
Description: (180-195) from US09041236.pep (11 of 45)
Perfect Score: 114
Sequence: 1 LYTSDVTVMQNPFQFIKA 16

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 27.109; Variance 33.102; scale 0.819

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
1	114	100.0	393	11	088371 SEMAPHORIN L (FRAGMENT	4.08e-14
2	114	100.0	666	4	075326 SEMAPHORIN L.	4.08e-14
3	89	78.1	633	14	064906 SIMILAR TO GENSUB ACC	1.81e-07
4	61	53.5	321	5	094652 IRON-SULFUR SUBUNIT OF	6.65e-01
5	60	52.6	389	10	080586 PUTATIVE N-MYRISTOYLTR	1.07e+00
6	60	52.6	577	10	082753 HYPOTHETICAL 64.1 KD P	1.07e+00
7	59	51.8	115	14	086516 ORF 2.	1.73e+00
8	59	51.8	334	11	088736 17-BETA-HYDROXYSTEROID	1.73e+00
9	59	51.8	457	10	080701 F8K4.15 PROTEIN.	1.73e+00
10	59	51.8	818	5	026059 PEROXINECTIN PRECURSOR	1.73e+00
11	59	51.8	1063	3	013983 HYPOTHETICAL 120.9 KD	1.73e+00
12	58	50.9	430	2	025052 ATP-DEPENDENT NUCLEASE	2.76e+00
13	58	50.9	439	10	P92962 PROLINE TRANSPORTER 2.	2.76e+00
14	58	50.9	1638	3	Q03921 HYPOTHETICAL 194.7 KD	2.76e+00
15	57	50.0	139	2	086460 HYPOTHETICAL 14.7 KD P	4.38e+00
16	57	50.0	945	2	026077 HELICASE.	4.38e+00
17	56	49.1	92	5	017456 HYPOTHETICAL 10.6 KD P	6.92e+00
18	56	49.1	334	11	062504 OVARIAN-SPECIFIC PROTE	6.92e+00
19	56	49.1	467	14	037369 NON-STRUCTURAL PROTEIN	6.92e+00
20	56	49.1	692	2	Q45772 OUTER MEMBRANE PROTEIN	6.92e+00

21	56	49.1	704	2	084360 HYPOTHETICAL 80.3 KD P	6.92e+00
22	56	49.1	1286	5	062018 B0391.6 PROTEIN.	6.92e+00
23	56	49.1	1589	5	045569 F54F11.2 PROTEIN.	6.92e+00
24	55	48.2	169	14	065225 ORF 13L.	1.09e+01
25	55	48.2	172	2	Q52850 VALINE START.	1.09e+01
26	55	48.2	298	5	Q27646 GLYCERALDEHYDEPHOSPHAT	1.09e+01
27	55	48.2	345	14	Q65182 PD345L.	1.09e+01
28	55	48.2	352	5	076864 EG:100G10.4 PROTEIN.	1.09e+01
29	55	48.2	409	2	Q25853 NADH-UBIQUINONE OXIDOR	1.09e+01
30	55	48.2	617	2	Q56794 XANTHOMONAS COMPESTRIS	1.09e+01
31	55	48.2	641	11	088501 CALPAIN-LIKE PROTEASE.	1.09e+01
32	55	48.2	641	11	Q35646 CALPAIN 6 (CALPAIN-LIK	1.09e+01
33	55	48.2	882	2	Q67594 ALGINATE SYNTHESIS-REL	1.09e+01
34	55	48.2	963	5	Q20878 F56D5.9 PROTEIN.	1.09e+01
35	55	48.2	1084	14	Q65353 ORF B.	1.09e+01
36	54	47.4	127	1	P95932 ORF C01034.	1.69e+01
37	54	47.4	186	6	Q29412 INTERFERON-BETA-1 PREC	1.69e+01
38	54	47.4	380	11	Q88321 ANTISECRETORY FACTOR.	1.69e+01
39	54	47.4	415	2	Q55407 HYPOTHETICAL 44.7 KD P	1.69e+01
40	54	47.4	482	14	P87550 NUCLEAR INCLUSION A (F	1.69e+01
41	54	47.4	666	2	Q54474 PENICILLIN-BINDING PROT	1.69e+01
42	54	47.4	666	2	P72530 PENICILLIN-BINDING PRO	1.69e+01
43	54	47.4	666	2	Q54781 PENICILLIN-BINDING PRO	1.69e+01
44	54	47.4	666	2	P72529 PENICILLIN-BINDING PRO	1.69e+01
45	54	47.4	751	14	Q39294 COUNTERPART OF HSV-1 G	1.69e+01

ALIGNMENTS

RESULT 1						
ID 088371	PRELIMINARY;	PRT;	393	AA.		
AC 088371;						
DT 01-NOV-1998	(TREMBLREL. 08, CREATED)					
DT 01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)					
DT 01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)					
DE SEMAPHORIN L (FRAGMENT).						
GN SEMAL.						
OS MUS MUSCULUS (MOUSE).						
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;						
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.						
RN [1]						
RP SEQUENCE FROM N.A.						
RX MEDLINE; 98389619.						
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;						
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA						
RT viruses."						
RL GENOMICS 51:340-350(1998).						
DR EMBL; AF030699; G3523117; -						
FT NON_TER 393 393						
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;						

Query Match 100.0%; Score 114; DB 11; Length 393;
Best Local Similarity 100.0%; Pred. No. 4.08e-14;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 210	LYTSDVTVMQNPFQFIKA 225
Qy 180	LYTSDVTVMQNPFQFIKA 195

RESULT 2						
ID 075326	PRELIMINARY;	PRT;	666	AA.		
AC 075326;						
DT 01-NOV-1998	(TREMBLREL. 08, CREATED)					
DT 01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)					
DT 01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)					
DE SEMAPHORIN L.						
GN SEMAL.						
OS HOMO SAPIENS (HUMAN).						
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;						
OC CATARRHINI; HOMINIDAE; HOMO.						
RN [1]						
RP SEQUENCE FROM N.A.						

```
RX MEDLINE: 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses.";
RL GENOMICS 51:340-350(1998).
DR EMBL: AF030698; G3523115; -.
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 114; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 4.08e-14;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 212 LYSDTVMQNPOFIKA 227
QY 180 LYSDTVMQNPOFIKA 195

RESULT 3
ID Q64906 PRELIMINARY; PRT; 653 AA.
AC Q64906;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALGELAPHINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RX MEDLINE: 97201573.
RA ENSSER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RX MEDLINE: 97404659.
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U18243; G1000717; -.
DR EMBL: AF005370; G2337970; -.
SQ SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match 78.1%; Score 89; DB 14; Length 653;
Best Local Similarity 68.8%; Pred. No. 1.81e-07;
Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 234 LYSDTAMHRPOFVQA 249
QY 180 LYSDTVMQNPOFIKA 195

RESULT 4
ID Q94652 PRELIMINARY; PRT; 321 AA.
AC Q94652;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE IRON-SULFUR SUBUNIT OF SUCCINATE DEHYDROGENASE.
OS PLASMIDIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1;
RA TAKEO S., KITA K.;
RT "Cloning of succinate dehydrogenase/fumarate reductase of Plasmodium
```

```
RT falciparum mitochondria.";
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: D86574; D1013807; -.
DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
DR PFAM; PF00037; fer4; 1.
KW IRON-SULFUR.
SQ SEQUENCE 321 AA; 37752 MW; 774D83B5 CRC32;

Query Match 53.5%; Score 61; DB 5; Length 321;
Best Local Similarity 33.3%; Pred. No. 6.65e-01;
Matches 5; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

Db 301 FSEDITIKHSQYIKS 315
QY 181 YTSDTVQNPOFIKA 195

RESULT 5
ID O80586 PRELIMINARY; PRT; 389 AA.
AC O80586;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE N-MYRISTOYLTRANSFERASE.
GN F6E13.30.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV COLUMBIA;
RA ROUNSLEY S.D., KAUL S., LIN X., KETCHUM K.A., CROSBY M.L.,
RA BRANSON R.C., SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RT "Arabidopsis thaliana chromosome II BAC F6E13 genomic sequence.";
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AC004005; G3212870; -.
KW TRANSFERASE.
SQ SEQUENCE 389 AA; 44450 MW; ED29F4BF CRC32;

Query Match 52.6%; Score 60; DB 10; Length 389;
Best Local Similarity 60.0%; Pred. No. 1.07e+00;
Matches 9; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Db 343 FYASD-VQNESFLK 356
QY 180 LYSDTVMQNPOFIK 194

RESULT 6
ID O82753 PRELIMINARY; PRT; 577 AA.
AC O82753;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 64.1 KD PROTEIN.
GN F7H19.230.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RA PETERS S.A., VAN STAVEREN M., DIRKSE W., STIEREMA W., BANCROFT I.,
RA MEWES H.W., MAYER K.F.X., SCHUELLER C., BEVAN M.;
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
EU ARABIDOPSIS SEQUENCING PROJECT;
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AL031018; E1310073; -.
```

KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 577 AA; 64063 MW; 78E7979E CRC32;

Query Match 52.6%; Score 60; DB 10; Length 577;
Best Local Similarity 42.9%; Pred. No. 1.07e+00;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 118 LFQNDPVMRPPFV 131
I : : : : :
QY 180 LYTSDTVMQNPQFI 193

RESULT 7
ID Q86516 PRELIMINARY; PRT; 115 AA.
AC Q86516;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ORF 2.
OS UNKNOWN.
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IDIR AGENT (INFECTIOUS DIARRHEA OF INFANT RATS);
RX MEDLINE; 94160571.
RA EIDEN J.J.;
RT "Expression and sequence analysis of gene 7 of the IDIR agent (group B rotavirus): similarity with NS53 of group A rotavirus.";
RL VIROLOGY 199:212-218(1994).
DR EMBL; U01164; G623030; -
SQ SEQUENCE 115 AA; 12845 MW; DA01E107 CRC32;

Query Match 51.8%; Score 59; DB 14; Length 115;
Best Local Similarity 53.3%; Pred. No. 1.73e+00;
Matches 8; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

Db 92 LYV-OPVMQSNPFK 105
I : : : : :
QY 180 LYTSDTVMQNPQFI 194

RESULT 8
ID Q88736 PRELIMINARY; PRT; 334 AA.
AC Q88736;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 17-BETA-HYDROXYSTEROID DEHYDROGENASE TYPE 7.
GN HSD17B7.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=MAMMARY GLAND;
RX MEDLINE; 98322544.
RA NIKELAINEN P.A.; PELTOKETO H.; VIHKO R.; VIHKO P.;
RT "Expression cloning of a novel estrogenic mouse
RT 17-beta-hydroxysteroid dehydrogenase/17-ketosteroid reductase
RT (m17HSD7), previously described as a prolactin receptor associated
RT protein (PRAP) in rat.";
RL MOL. ENDOCRINOL. 12:1048-1059(1998).
DR EMBL; Y15733; E1311087; -
SQ SEQUENCE 334 AA; 37317 MW; 3625B946 CRC32;

Query Match 51.8%; Score 59; DB 11; Length 334;
Best Local Similarity 47.1%; Pred. No. 1.73e+00;
Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

Db 87 LYLNAGILPNQFNKA 103
I : : : : :
QY 180 LYTSDTVMQNPQFI-1KA 195

RESULT 9
ID Q80701 PRELIMINARY; PRT; 457 AA.
AC Q80701;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F8K4.15 PROTEIN.
GN F8K4.15.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA VYSOTSKAIA V.S.; SCHWARTZ J.R.; TORIUMI M.; KWAN A.; YU G.; OJI O.;
RA LIU S.; LI J.; ARAUJO R.; AU M.; BRENDEN V.; BUEHLER E.; CONWAY A.B.;
RA CONWAY A.R.; DEWAR K.; FENG J.; KIM C.; KURTZ D.; LI Y.; PALM C.J.;
RA SHINN P.; SUN H.; DAVIS R.W.; ECKER J.R.; FEDERSPIEL N.A.;
RA THEOLOGIS A.;
RT "Arabidopsis thaliana chromosome 1 BAC F8K4 sequence.";
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA THEOLOGIS;
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]

Query Match 51.8%; Score 59; DB 10; Length 457;
Best Local Similarity 46.2%; Pred. No. 1.73e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 358 YTAETVKKKTEFI 370
I : : : : :
QY 181 YTSDTVQNPQFI 193

RESULT 10
ID Q26059 PRELIMINARY; PRT; 818 AA.
AC Q26059;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PEROXINECTIN PRECURSOR.
OS PACIFASTACUS LENIUSCULUS (SIGNAL CRAYFISH).
OC EUKARYOTA; METAZOA; ARTHROPODA; CRUSTACEA; MALACOSTRACA;
OC EUMALACOSTRACA; EUCARIDA; DECAPODA; PLEOCYENATA; ASTACIDEA; ASTACOIDEA;
OC ASTACIDAE; PACIFASTACUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96074593.
RA JOHANSSON M.W.; LIND M.I.; HOLMBLAD T.; THOERNQVIST P.O.;
RA SODERHALL K.;
RT "Peroxinectin, a novel cell adhesion protein from crayfish blood.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 216:1079-1087(1995).

```
DR EMBL: X91409; E199227; -.
DR PFAM: PF00141; peroxidase; 2.
KW SIGNAL.
FT SIGNAL. 1 18 POTENTIAL.
FT CHAIN 19 818 POTENTIAL.
SQ SEQUENCE 818 AA; 89315 MW; 777957A9 CRC32;

Query Match 51.8%; Score 59; DB 5; Length 818;
Best Local Similarity 64.3%; Pred. No. 1.73e+00;
Matches 9; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 314 LFTS-SVMQWQFI 326
QY 180 LYTSDTVMQNPQFI 193
||| :||| :|||
RESULT 11
ID O13983 PRELIMINARY; PRT; 1063 AA.
AC O13983; O42856;
DT 01-JAN-1999 (TREMBLREL. 09, CREATED)
DT 01-JAN-1999 (TREMBLREL. 09, LAST SEQUENCE UPDATE)
DE 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 120.9 KD PROTEIN C26H5.O1C IN CHROMOSOME 1.
GN SPAC26H5.O1C OR SPAC3A1.18C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO THE DNA HELICASES.
DR EMBL: 299126; E1198268; -.
DR EMBL: AL021813; E1250592; -.
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 167 170 POLY-ARG.
SQ SEQUENCE 1063 AA; 120969 MW; 98C8ADD5 CRC32;

Query Match 51.8%; Score 59; DB 3; Length 1063;
Best Local Similarity 42.9%; Pred. No. 1.73e+00;
Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 139 LYTGSTVSQSSEYV 152
QY 180 LYTSDTVMQNPQFI 193
||| :||| :|||
RESULT 12
ID O25052 PRELIMINARY; PRT; 430 AA.
AC O25052;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ATP-DEPENDENT NUCLEASE (ADDB).
GN HP0275.
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC HELICOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695;
RX MEDLINE: 97394467.
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKIY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;

RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori [published erratum appears in Nature 1997 Sep
RL NATURE 388:539-547(1997).
DR EMBL: AE000546; G2313366; -.
DR TIGR: HP0275; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 430 AA; 49684 MW; 1377C500 CRC32;

Query Match 50.9%; Score 58; DB 2; Length 430;
Best Local Similarity 46.2%; Pred. No. 2.76e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 223 LYEKNPVQNAQF 235
QY 180 LYTSDTVMQNPQF 192
||| :||| :|||
RESULT 13
ID P92962 PRELIMINARY; PRT; 439 AA.
AC P92962;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE PROLINE TRANSPORTER 2.
GN PROT2.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. LANDSBERG; TISSUE=LEAF;
RX MEDLINE: 96373203.
RA RENTSCH D., HIRNER B., SCHMELZER E., FROMMER W.B.;
RT "Salt stress-induced proline transporters and salt stress-repressed
RT broad specificity amino acid permeases identified by suppression of a
RT yeast amino acid permease-targeting mutant.";
RL PLANT CELL 8:1437-1446(1996).
DR EMBL: X95738; E225741; -.
DR MENDEL; 16598; ARATH; 2547; m16598.
SQ SEQUENCE 439 AA; 48085 MW; 2CA89AE5 CRC32;

Query Match 50.9%; Score 58; DB 10; Length 439;
Best Local Similarity 42.9%; Pred. No. 2.76e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 142 LFRDSDLMKLPHF 155
QY 180 LYTSDTVMQNPQFI 193
||| :||| :|||
RESULT 14
ID Q03921 PRELIMINARY; PRT; 1698 AA.
AC Q03921;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 194.7 KD PROTEIN.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA OLIVER K., HARRIS D.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RC BARRELL B., RAJANDREAM M.A.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: Z48179; G665674; -.

```


KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1698 AA; 194685 MW; 38B43DAC CRC32;
Query Match 50.9%; Score 58; DB 3; Length 1698;
Best Local Similarity 43.8%; Pred. No. 2.76e+00;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
Db 436 LYSSEKFKQNEQIMKT 451
::: :|| :|| :||
Qy 180 LYSDTVMQNPQFIKA 195
RESULT 15
ID O86460 PRELIMINARY; PRT; 139 AA.
AC O86460;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 14.7 KD PROTEIN (FRAGMENT).
OS RHIZOBIUM LEGUMINOSARUM.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; RHIZOBIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8401(PRLJII);
RA STEVENS J.B., JOHNSTON A.W.B.;
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AJ007906; E1315920; -
KW HYPOTHETICAL PROTEIN.
FT NON_TER 139 139
SQ SEQUENCE 139 AA; 14704 MW; DC42209F CRC32;
Query Match 50.0%; Score 57; DB 2; Length 139;
Best Local Similarity 30.8%; Pred. No. 4.38e+00;
Matches 4; Conservative 8; Mismatches 1; Indels 0; Gaps 0;
Db 19 FAADAVVQDPAYV 31
::: :|| :|| :||
Qy 181 YTSDTVMPQNPQFI 193

Search completed: Thu Jul 8 18:27:43 1999
Job time : 19 secs.

THIS PAGE BLANK (USPTO)

W P E R E A (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:30:33 1999; Maspar time 11.62 Seconds
Tabular output not generated. 14.638 Million cell updates/sec

Title: >US-09-041-236-2
Description: (201-208) from US09041236.pap (12 of 45)
Perfect Score: 55
Sequence: 1 DQAYDDKI 8

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 15.523; Variance 42.966; scale 0.361

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match	Length DB ID	Description	
1	45	81.8	1206 30 W47080	Tomato Mi resistance	1.23e+02
2	45	81.8	1206 30 W3583	Wild tomato Mi resist	1.23e+02
3	45	81.8	1257 30 W3582	Wild tomato Mi resist	1.23e+02
4	44	80.0	313 33 W5044	Thuja pinoselinol-lar	1.60e+02
5	44	80.0	349 19 W06416	Phosphotriesterase-re	1.60e+02
6	43	78.2	415 28 W19005	Feline herpes virus t	2.09e+02
7	42	76.4	208 39 W73628	Human secreted protei	2.71e+02
8	42	76.4	248 22 W02610	H. pylori secreted or	2.71e+02
9	42	76.4	248 22 W20433	H. pylori secreted or	2.71e+02
10	41	74.5	337 15 P81191	Biotin synthetase (Bi	3.51e+02
11	41	74.5	1607 35 W50897	Mouse laminin G1 chai	3.51e+02
12	41	74.5	1609 35 W50898	Human laminin G1 chai	3.51e+02
13	40	72.7	113 31 W27988	Amino acid sequence o	4.54e+02
14	40	72.7	146 34 W77589	Elongation factor Tu	4.54e+02
15	40	72.7	155 27 W24297	Staphylococcus aureus	4.54e+02
16	40	72.7	155 33 W57836	Spo-rel protein seque	4.54e+02

17	40	72.7	155 31 W27962	Amino acid sequence o	4.54e+02
18	40	72.7	251 29 W47421	Bacillus subtilis pre	4.54e+02
19	40	72.7	312 32 W5047	Thuja pinoselinol-lar	4.54e+02
20	40	72.7	488 35 W54569	Potato solanidine UPP	4.54e+02
21	40	72.7	736 33 W57835	Spo-rel protein seque	4.54e+02
22	39	70.9	110 1 R06103	Fowl Pox Virus protei	5.85e+02
23	39	70.9	177 2 P93650	Amino acid sequence o	5.85e+02
24	39	70.9	177 1 R06440	MG-5 antigen.	5.85e+02
25	39	70.9	218 6 R34483	atp-integrase region	5.85e+02
26	39	70.9	276 3 R13467	cc protein.	5.85e+02
27	39	70.9	285 1 P82114	Peptide corresp. to f	5.85e+02
28	39	70.9	325 10 R52028	Protein with Oxetanoc	5.85e+02
29	39	70.9	344 7 R34579	Phase I5 integrase pr	5.85e+02
30	39	70.9	344 4 R20980	putative Mycobacteria	5.85e+02
31	39	70.9	385 2 R10652	Cephalosporin antibio	5.85e+02
32	39	70.9	432 31 W40594	E. coli ADSS protein.	5.85e+02
33	39	70.9	501 35 W64782	D. multivorans PCE-De	5.85e+02
34	39	70.9	524 1 P82585	Modified tissue plas	5.85e+02
35	39	70.9	562 3 P20285	Sequence encoded by h	5.85e+02
36	39	70.9	577 22 W20774	H. pylori flagella-as	5.85e+02
37	39	70.9	652 16 R80037	Bacillus stearothermo	5.85e+02
38	39	70.9	652 17 R96109	Starch-branching-enz	5.85e+02
39	39	70.9	829 16 R86578	Autotaxin derived fro	5.85e+02
40	39	70.9	849 16 R86595	N-tera 2D1 autotaxin	5.85e+02
41	39	70.9	915 16 R86596	A2058 autotaxin prote	5.85e+02
42	39	70.9	2050 38 W73499	Von Willebrand factor	5.85e+02
43	39	70.9	2182 23 W22476	Plasmodium var-1.	5.85e+02
44	39	70.9	2813 3 P60053	Sequence of von Wille	5.85e+02
45	39	70.9	2813 3 P60462	Sequence of human von	5.85e+02

ALIGNMENTS

RESULT 1
ID W47080 standard; Protein; 1206 AA.
AC W47080;
DT 20-JUL-1998 (first entry)
DE Tomato Mi resistance gene product.
KW Mi resistance gene; tomato; nematode resistance;
KM Meloidogyne incognita; crop protection; transgenic plant.
OS Lycopersicon esculentum.
PN EP-823481-A1.
PD 11-FEB-1998.
PF 09-AUG-1996; 401764.
PR 09-AUG-1996; EP-401764.
PA (KEYG-) KEYGENE NV.
PI Simons G, Vos P, Wijnbrandi J, Zabeau M;
DR WPI: 98-112270/11.
DR N-PSDB; V13935.
PT Tomato Mi resistance gene - for producing nematode-resistant
transgenic plants
PS Claim 39; Fig 5A-D; 48pp; English.
CC This polypeptide is encoded by the tomato Mi resistance gene (see
V13935) that confers resistance against nematodes. The invention
relates to genes, such as the Mi resistance gene, capable of
confering nematode resistance, and to genetically transformed
plants which are resistant to nematodes, especially Meloidogyne
incognita. In addition to expression in plants, the Mi resistance
gene product can be expressed in bacterial or yeast cells.
CC Antibodies raised against the polypeptide can be used to detect the
presence of the Mi resistance gene product.
SQ Sequence 1206 AA;

Query Match 81.8%; Score 45; DB 30; Length 1206;
Best Local Similarity 85.7%; Pred. No. 1.23e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 531 dqgyddk 537

Qy 201 DQAYDDK 207

RESULT 2

ID W53583 standard; Protein: 1206 AA.
AC W53583;
DE 07-JUL-1998 (first entry)
DE Wild tomato Mi resistance gene open reading frame 2 product.
KW AFLP marker PM14; root knot nematode; Mi; resistance gene;
KW wild tomato; pathogen resistance; pest resistance; aphid;
KW Meloidogyne incognita.
OS Lycopersicon peruvianum.
PN W09806750-AA2.
PD 19-FEB-1998.
PE 08-AUG-1997; E04340.
PR 16-MAY-1997; EP-401101.
PR 09-AUG-1996; EP-401764.
PA (KEYG-) KEYGENE NV.
PI Simons G, Vos P, Wijbrandi J, Zabeau M;
DR WPI: 98-159460/14.
DR N-PSDB: V16457.
PT New isolated Meloidogyne incognita resistance gene - used to transform plants to provide resistance to pathogens or pests, particularly nematodes or aphids.
PS Disclosure; Fig 7B; 79pp; English.
CC The present sequence is encoded by an approximately 9.9 kb sequence located around the wild tomato, Lycopersicon peruvianum (PI 128657), AFLP marker PM14, which comprises the root knot nematode Meloidogyne incognita (Mi) resistance gene. The gene can be used to transform plants to provide resistance to plant pathogens or pests, particularly nematodes, e.g. root-knot nematodes, or aphids.
SQ Sequence 1206 AA;

Query Match 81.8%; Score 45; DB 30; Length 1206;
Best Local Similarity 85.7%; Pred. No. 1.23e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 531 dqgyddk 537
|:|:|:|
QY 201 DQAYDDK 207

RESULT 3
ID W53582 standard; Protein: 1257 AA.
AC W53582;
DE 07-JUL-1998 (first entry)
DE Wild tomato Mi resistance gene open reading frame 1 product.
KW AFLP marker PM14; root knot nematode; Mi; resistance gene;
KW wild tomato; pathogen resistance; pest resistance; aphid;
KW Meloidogyne incognita.
OS Lycopersicon peruvianum.
PN W09806750-AA2.
PD 19-FEB-1998.
PE 08-AUG-1997; E04340.
PR 16-MAY-1997; EP-401101.
PR 09-AUG-1996; EP-401764.
PA (KEYG-) KEYGENE NV.
PI Simons G, Vos P, Wijbrandi J, Zabeau M;
DR WPI: 98-159460/14.
DR N-PSDB: V16457.
PT New isolated Meloidogyne incognita resistance gene - used to transform plants to provide resistance to pathogens or pests, particularly nematodes or aphids.
PS Claim 43; Fig 7A; 79pp; English.
CC The present sequence is encoded by an approximately 9.9 kb sequence located around the wild tomato, Lycopersicon peruvianum (PI 128657), AFLP marker PM14, which comprises the root knot nematode Meloidogyne incognita (Mi) resistance gene. The gene can be used to transform plants to provide resistance to plant pathogens or pests, particularly nematodes, e.g. root-knot nematodes, or aphids.
SQ Sequence 1257 AA;

Query Match 81.8%; Score 45; DB 30; Length 1257;
Best Local Similarity 85.7%; Pred. No. 1.23e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 582 dqgyddk 588
|:|:|:|
QY 201 DQAYDDK 207

RESULT 4
ID W65044 standard; Protein: 313 AA.
AC W65044;
DE 28-SEP-1998 (first entry)
DE Thuja pinoresinol-lariciresinol reductase PLR-Tp1.
KW pinoresinol-lariciresinol reductase; dirigent protein; lignan;
KW transgenic plant; anticancer; cytostatic; antiviral; virucide;
KW antibiotic; antioxidant; antifeedant.
OS Thuja plicata.
PN W09820113-AA1.
PD 14-MAY-1998.
PE 07-NOV-1997; U20391.
PR 31-JUL-1997; US-054380.
PR 08-NOV-1996; US-030522.
PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
PI Davin LB, Dinkovakostova AT, Fujita M, Gang DR,
PI Lewis NG, Sarkanen S;
DR WPI: 98-286929/25.
DR N-PSDB: V35182.
PT New plant-derived dirigent proteins and pinoresinol-lariciresinol reductases - and related nucleic acid, vectors and transformants, used for stereospecific production of lignans, useful e.g. as anticancer or antiviral agents
PS Claim 17; Page 122-123; 148pp; English.
CC PLR-Tp1 comprises a (-)-pinoresinol/(-)-lariciresinol reductase (P/LR) of Thuja plicata. P/LR enzymes catalyze the conversion of pinoresinol to lariciresinol and then to secoisolariciresinol. 4 Isoforms (see W65044-47) of T. plicata P/LR have been identified; the other 3 isoforms are (+)/(-)-P/LR. Further P/LR enzymes have been obtained from Forsythia intermedia (see W65038-43) and Tsuga heterophylla (see W65048-49). The isolation of cDNAs encoding P/LR (see V35175-80 and V35182-87) and dirigent proteins (see V35160-71) permits the development of an efficient expression system for those enzymes, provides useful tools for examining the developmental regulation of lignan biosynthesis and permits the isolation of related sequences. It also allows the transformation of a wide range of organisms, including plants, in order to modify lignan biosynthesis. Optically pure lignans may have e.g. anticancer, antiviral, antioxidant, antibiotic or antifeedant activity.
SQ Sequence 313 AA;

Query Match 80.0%; Score 44; DB 33; Length 313;
Best Local Similarity 50.0%; Pred. No. 1.60e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 260 dksyeeki 267
|:|:|:|
QY 201 DQAYDDK 208

RESULT 5
ID W06416 standard; protein: 349 AA.
AC W06416;
DE 28-JAN-1997 (first entry)
DE Phosphotriesterase-related protein.
KW Phosphotriesterase-related protein; mouse; PRP; recessive cpk; murine;
KW polycystic kidney disease; PKD; acute renal injury; autosomal dominant;
KW autosomal recessive; fluid-filled cyst; nephron; collecting duct; kidney;
KW renal parenchyma; renal failure; mammal; diagnostic marker; therapy;
KW acute organophosphate toxicity; phosphotriesterase inhibitor;
KW anti-cancer drug enhancer; anti-HIV drug enhancer.
OS Mus musculus.
PN US5552313-A.
PD 03-SEP-1996.
PR 21-NOV-1994; 343027.
PR 21-NOV-1994; US-343027.
PA (UNIV) UNIV KANSAS.

PI Calvet JP, Hou X, Magenheimer BS, Maser RL;
 DR WPI: 96-412067/41.
 DR N-PSDB; T43210.
 PT Mouse phospho:tri:esterase-related protein DNA - used to develop
 PT prods. for diagnosis of poly:cystic kidney disease or acute renal
 PT failure
 PS Claim 1: Column 25-28; 23pp; English.
 CC This sequence represents the mouse phosphotriesterase related protein
 CC (MPRP). The MPRP sequence has abnormal underexpression in the recessive
 CC cpk murine model of a polycystic kidney disease (PKD), and has decreased
 CC expression following acute renal injury. PKD is a common disease
 CC (affecting 1 in 500-1000 individuals) and can be inherited as an
 CC autosomal dominant or as an autosomal recessive. PKD is characterised
 CC by the development of innumerable, large, fluid-filled cysts from the
 CC nephrons and collecting ducts of affected kidneys. It is thought that
 CC enlargement of the cysts interferes with functioning of the normal renal
 CC parenchyma, which eventually leads to renal failure. The MPRP, or other
 CC similar mammalian PRPs can be used as diagnostic markers for PKD and
 CC acute renal failure. The PRPs can also be used in therapeutic and
 CC protective treatments, such as for acute organophosphate toxicity. They
 CC can also be used to develop products to enhance the effectiveness of
 CC other types of therapy, e.g. phosphotriesterase inhibitors to enhance the
 CC effectiveness of certain anti-cancer, or anti-HIV drugs.
 SQ Sequence 349 AA;

Query Match 80.08; Score 44; DB 19; Length 349;
 Best Local Similarity 50.08; Pred. No. 1.60e+02;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 286 degyedik 293

l::l::l

QY 201 DQAYDDKI 208

RESULT 6
 ID W19005 standard; Protein; 415 AA.
 AC W19005;
 DT 05-MAY-1998 (first entry)
 DE Feline herpes virus type 1 ORF5-encoded protein.
 CC Feline herpes virus type 1; open reading frame; ORF; vector; vaccine;
 KW feline infectious peritonitis virus; FIPV; cat; immunisation;
 KW rhinotracheitis.
 OS Feline herpesvirus.
 PN W09720059-AL.
 PD 05-JUN-1997.
 PF 19-NOV-1996; F01830.
 PR 30-NOV-1995; FR-014450.
 PA (INMR) RHONE MERIEUX SA.
 PI Audonnet JCF, Baudu PGN, Riviere MAE;
 DR WPI: 97-310613/28.
 DR N-PSDB; T69857.
 PT Live recombinant vaccine based on feline herpes virus - has
 PT antigen-encoding sequence inserted in open reading frame 2 or 5,
 PT particularly for protection against feline infectious peritonitis
 PT virus
 PS Claim 18: Fig 1; 60pp; French.
 CC This sequence represent the protein encoded by ORF5 from the feline
 CC herpes virus type 1 (FHV-1) from strain CO. The nucleotide sequence
 CC is used as a vector to generate a live recombinant vaccine, in which a
 CC polypeptide coding sequence (especially a gene taken from the feline
 CC infectious peritonitis virus (FIPV)) is inserted into open reading
 CC frames 5 and/or 2. Vaccines comprising the vector are used to protect
 CC cats, specifically against FIPV. The vaccine is attenuated but retains
 CC a good capacity to replicate in vivo and still protects against
 CC infectious rhinotracheitis (caused by FHV).
 SQ Sequence 415 AA;

Query Match 78.28; Score 43; DB 28; Length 415;
 Best Local Similarity 50.08; Pred. No. 2.09e+02;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 257 edvyedik 264

l::l::l

QY 201 DQAYDDKI 208

RESULT 7
 ID W73628 standard; protein; 208 AA.
 AC W73628;
 DT 23-MAR-1999 (first entry)
 DE Human secreted protein clone bv227_1.
 CC Secreted protein; human; nutritional supplements; immune stimulant;
 KW immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
 KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
 KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
 KW cadherin/tumour invasion suppressor.
 OS Homo sapiens.
 PN W09855614-AZ.
 PD 10-DEC-1998.
 PF 01-JUN-1998; U11210.
 PR 29-MAY-1998; US-087255.
 PR 04-JUN-1997; US-868696.
 PR 04-JUN-1997; US-868697.
 PR 04-JUN-1997; US-868698.
 PR 04-JUN-1997; US-868898.
 PR 04-JUN-1997; US-868899.
 PR 04-JUN-1997; US-868900.
 PR 04-JUN-1997; US-869191.
 PR 04-JUN-1997; US-869192.
 PR 04-JUN-1997; US-869193.
 PR 04-JUN-1997; US-869194.
 PA (GEM) GENETICS INST INC.
 PI Agostino MJ, Fecthel K, Howes SH, Jacobs K, Lavallie ER,
 PI McCoy JM, Racine LA, Spaulding V, Treacy M;
 DR WPI: 99-059912/05.
 DR N-PSDB; V55746.

PT New polynucleotides encoding secreted human proteins - derived from
 PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
 PT retina, adult placenta or adult uterus cDNA libraries
 PS Claim 25: Page 89; 127pp; English.
 CC This sequence represents a human secreted protein of the invention.
 CC The DNA encoding this sequence was isolated from a human adult brain
 CC cDNA library, and was designated clone bv227_1. The DNAs and proteins
 CC are predicted to have biological activities which would make them
 CC suitable for treating, preventing or ameliorating medical conditions in
 CC humans and animals, although no supporting data is given. Suggested
 CC activities include nutritional sources or supplements, immune
 CC stimulating or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The DNAs
 CC are also stated to be useful for gene therapy. A host cell transfected
 CC with the DNA, or its subfragments and variants is useful for recombinant
 CC production of the human secreted protein clones.
 SQ Sequence 208 AA;

Query Match 76.48; Score 42; DB 39; Length 208;
 Best Local Similarity 71.48; Pred. No. 2.71e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 77 qtyddkl 83

l::l::l

QY 202 QAYDDKI 208

RESULT 8
 ID W20610 standard; protein; 248 AA.
 AC W20610;
 DT 14-JUL-1997 (first entry)
 DE H. pylori secreted or periplasmic protein, O1xe2171orf12.
 KW Vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
 KW secreted; periplasmic.
 OS Helicobacter pylori.

PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR.) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR N-PSDB; T67863.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 72; Pages 1038-1039; 1481pp; English.
 CC The present sequence is a Helicobacter pylori secreted or
 CC periplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat
 CC H. pylori infection or to identify H. pylori polypeptide binding
 CC compounds, useful as potential H. pylori life cycle activators or
 CC inhibitors, the genomic sequence of H. pylori (ATCC 55679) was
 CC determined from overlapping contigs generated by mechanically
 CC shearing the bacterial DNA. The sequences were analysed for ORF of
 CC at least 180 nucleotides, and the predicted coding regions defined
 CC by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from
 CC various ORF were analysed for significant homology to other known
 CC or exported membrane proteins. Having identified and determined
 CC the sequences of interest, particular regions can be isolated from
 CC H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 248 AA;

Query Match 76.4%; Score 42; DB 22; Length 248;
 Best Local Similarity 75.0%; Pred. No. 2.71e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 48 dgaedll 55
 |||||
 QY 201 DQAYDDKI 208

RESULT 9
 ID W20433 standard; protein; 248 AA.
 AC W20433.
 DT 14-JUL-1997 (first entry)
 DE H. pylori secreted or periplasmic protein, 34179577.aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR.) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR N-PSDB; T67607.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 72; Pages 610; 1481pp; English.
 CC The present sequence is a Helicobacter pylori secreted or periplasmic
 CC protein. The protein may be used in a vaccine to prevent or treat
 CC H. pylori infection or to identify H. pylori polypeptide binding
 CC compounds, useful as potential H. pylori life cycle activators or
 CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was
 CC determined from overlapping contigs generated by mechanically shearing
 CC the bacterial DNA. The sequences were analysed for ORF of at least 180
 CC nucleotides, and the predicted coding regions defined by computer
 CC evaluation. To identify likely H. pylori antigens for vaccine
 CC development, the amino acid sequences predicted from various ORF were
 CC analysed for significant homology to other known or exported membrane

CC proteins. Having identified and determined the sequences of interest,
 CC particular regions can be isolated from H. pylori by PCR amplification
 CC for recombinant polypeptide production, e.g. in E. coli hosts.
 SQ Sequence 248 AA;

Query Match 76.4%; Score 42; DB 22; Length 248;
 Best Local Similarity 75.0%; Pred. No. 2.71e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 48 dgaedll 55
 |||||
 QY 201 DQAYDDKI 208

RESULT 10
 ID P81191 standard; protein; 337 AA.
 AC P81191.
 DT 04-DEC-1990 (first entry)
 DE Biotin synthetase (Bio B) of B.sphaericus IFO 3525.
 KW biotin biosynthesis; fermentative biotin production; plasmid pTG1400;
 KW biotin synthetase; Bio B; seborrhalic dermatitis treatment.
 OS Bacillus sphaericus.
 PN EP-266240-A.
 PD 04-MAY-1988.
 PF 28-SEP-1987; 402157.
 PR 30-SEP-1986; FR-013603.
 PR 18-MAY-1987; FR-006916.
 PA (TRAN-) Transgene SA.
 PI Gloekler R, Speck D, Lemoine Y.;
 DR WPI: 88-121306/18.
 DR N-PSDB; N81546.
 PT New DNA sequences encoding enzymes involved in biotin biosynthesis -
 PT isolated from Bacillus sphaericus, and transformants useful in
 PT fermentative biotin production.
 PS Disclosure; p; French.
 CC DNA was isolated from B.sphaericus IFO 3525, cut with HindIII and
 CC the fragments sub-cloned into pBR322. The recombinant plasmids were
 CC used to transform E.coli bio- mutants. Plasmid pTG1400 was present
 CC in one clone (E.coli C600; CNCM I-608) which showed complementation
 CC for bio A and bio D. The plasmid contains a 4.3kb insert which
 CC contains 4 long open reading frames (LORFs) including that which
 CC encodes bio B. This enzyme is well characterised.
 CC A second clone was isolated which contained DNA encoding bio F and
 CC bio C. The insert present in the plasmid of this clone and the 4.3kb
 CC insert from pTG1400 were cloned together in pBR322 to produce a
 CC plasmid (pTG1440) which complements biotin auxotrophy in a bio delta
 CC FCD mutant.
 CC Transformed cells are used to produce biotin by culture in a medium
 CC contg pimelic acid and/or biotin vitamer. Co-culture of two
 CC transformants is preferred where one synthesises the vitamer from
 CC pimelic acid and the other converts the vitamer to biotin.
 CC See also N81551.
 SQ Sequence 337 AA;

Query Match 74.5%; Score 41; DB 1; Length 337;
 Best Local Similarity 42.9%; Pred. No. 3.51e+02;
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 185 qtyedrv 191
 |::|::
 QY 202 QAYDDKI 208

RESULT 11
 ID W50897 standard; Protein; 1607 AA.
 AC W50897.
 DT 07-DEC-1998 (first entry)
 DE Mouse laminin G1 chain.
 KW Laminin; mouse; beta-amyloid; amyloidosis; Alzheimer's disease;
 KW Down's syndrome; hereditary cerebral haemorrhage; inflammation;
 KW malignancy; Familial Mediterranean Fever; multiple myeloma;
 KW type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;
 KW Gertsmann-Straussler syndrome; kuru; scrapie; haemodialysis;

KW carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
 KW Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;
 KW therapy.
 OS Mus sp.
 PN WO9815179-A1.
 PD 16-APR-1998.
 PF 08-OCT-1997; U18145.
 PF 08-OCT-1996; US-027981.
 PA (UNIW) UNIV WASHINGTON.
 PI Castillo G, Snow AD;
 PR 08-OCT-1996; US-027981.
 PA (UNIW) UNIV WASHINGTON.
 PI Castillo G, Snow AD;
 DR WPI; 98-240534/21.
 PT Use of laminin and fragments - for developing products for use in
 PT the diagnosis and treatment of amyloid disease, e.g. Alzheimer's
 PT disease or CJD
 PS Claim 15; Page 102-105; 132pp; English.
 CC This is the amino acid sequence of the mouse laminin G1 chain. The
 CC primary object of the invention is to use laminin, laminin-derived
 CC protein fragments and/or laminin-derived polypeptides as potent
 CC inhibitors of amyloid formation, deposition, accumulation and/or
 CC persistence in Alzheimer's disease and other amyloidosis. The
 CC laminin products (see W50888-98) may include mouse or human laminin
 CC A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (merosin),
 CC laminin G1 chain, the globular repeats of the laminin A1 chain and
 CC the beta-amyloid binding domain of the laminin A chain. A claimed
 CC method for treating an amyloid disease comprises administering a
 CC polypeptide having a conformational similarity to a fragment of a
 CC laminin protein. A method for diagnosing an amyloid disease
 CC involves determining levels of laminin in a sample. Production
 CC of laminin or its fourth globular repeat in vivo provides a method
 CC for in vivo inhibition of beta-amyloid amyloidosis. The products
 CC and methods can be used for the diagnosis, prognosis, monitoring
 CC and treatment of amyloidosis such as Alzheimer's disease, Down's
 CC syndrome and hereditary cerebral haemorrhage with amyloidosis of
 CC the Dutch type (where the specific amyloid is the beta-amyloid
 CC protein), the amyloidosis associated with chronic inflammation,
 CC various forms of malignancy and Familial Mediterranean Fever (AA
 CC amyloid or inflammation-association amyloidosis), the amyloidosis
 CC associated with multiple myeloma and other B-cell abnormalities
 CC (AL amyloid), the amyloidosis associated with type II diabetes
 CC (amylin or islet amyloid), the amyloidosis associated with prion
 CC diseases including Creutzfeldt-Jacob disease, Gerstmann-Straussler
 CC syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis
 CC associated with long-term haemodialysis and carpal tunnel syndrome
 CC (beta 2-microglobulin amyloid), the amyloidosis associated with
 CC senile cardiac amyloid and Familial Amyloidotic Polynuropathy
 CC (prealbumin or transthyretin amyloid), and the amyloidosis
 CC associated with endocrine tumours such as medullary carcinoma of
 CC the thyroid (variant of procalcitonin).
 CC Sequence 1607 AA;

Query Match 74.5%; Score 41; DB 35; Length 1607;
 Best Local Similarity 50.0%; Pred. No. 3.51e+02;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 1066 dqafedrl 1073
 |||:::
 QY 201 DQAYDDKI 208

RESULT 12
 ID W50898 standard; Protein; 1609 AA.

AC W50898;
 DE 07-DEC-1998 (first entry)
 DE Human laminin G1 chain.
 KW Laminin; human; beta-amyloid; amyloidosis; Alzheimer's disease;
 KW Down's syndrome; hereditary cerebral haemorrhage; inflammation;
 KW malignancy; Familial Mediterranean Fever; multiple myeloma;
 KW type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;
 KW Gerstmann-Straussler syndrome; kuru; scrapie; haemodialysis;
 KW carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
 KW Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;
 KW therapy.
 OS Homo sapiens.

PN WO9815179-A1.
 PD 16-APR-1998.
 PF 08-OCT-1997; U18145.
 PF 08-OCT-1996; US-027981.
 PA (UNIW) UNIV WASHINGTON.
 PI Castillo G, Snow AD;
 PR 08-OCT-1996; US-027981.
 PA (UNIW) UNIV WASHINGTON.
 PI Castillo G, Snow AD;
 DR WPI; 98-240534/21.
 PT Use of laminin and fragments - for developing products for use in
 PT the diagnosis and treatment of amyloid disease, e.g. Alzheimer's
 PT disease or CJD
 PS Claim 15; Page 106-109; 132pp; English.
 CC This is the amino acid sequence of the human laminin G1 chain. The
 CC primary object of the invention is to use laminin, laminin-derived
 CC protein fragments and/or laminin-derived polypeptides as potent
 CC inhibitors of amyloid formation, deposition, accumulation and/or
 CC persistence in Alzheimer's disease and other amyloidosis. The
 CC laminin products (see W50888-98) may include mouse or human laminin
 CC A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (merosin),
 CC laminin G1 chain, the globular repeats of the laminin A1 chain and
 CC the beta-amyloid binding domain of the laminin A chain. A claimed
 CC method for treating an amyloid disease comprises administering a
 CC polypeptide having a conformational similarity to a fragment of a
 CC laminin protein. A method for diagnosing an amyloid disease
 CC involves determining levels of laminin in a sample. Production
 CC of laminin or its fourth globular repeat in vivo provides a method
 CC for in vivo inhibition of beta-amyloid amyloidosis. The products
 CC and methods can be used for the diagnosis, prognosis, monitoring
 CC and treatment of amyloidosis such as Alzheimer's disease, Down's
 CC syndrome and hereditary cerebral haemorrhage with amyloidosis of
 CC the Dutch type (where the specific amyloid is the beta-amyloid
 CC protein), the amyloidosis associated with chronic inflammation,
 CC various forms of malignancy and Familial Mediterranean Fever (AA
 CC amyloid or inflammation-association amyloidosis), the amyloidosis
 CC associated with multiple myeloma and other B-cell abnormalities
 CC (AL amyloid), the amyloidosis associated with type II diabetes
 CC (amylin or islet amyloid), the amyloidosis associated with prion
 CC diseases including Creutzfeldt-Jacob disease, Gerstmann-Straussler
 CC syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis
 CC associated with long-term haemodialysis and carpal tunnel syndrome
 CC (beta 2-microglobulin amyloid), the amyloidosis associated with
 CC senile cardiac amyloid and Familial Amyloidotic Polynuropathy
 CC (prealbumin or transthyretin amyloid), and the amyloidosis
 CC associated with endocrine tumours such as medullary carcinoma of
 CC the thyroid (variant of procalcitonin).
 CC Sequence 1609 AA;

Query Match 74.5%; Score 41; DB 35; Length 1609;
 Best Local Similarity 50.0%; Pred. No. 3.51e+02;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 1068 dqafedrl 1075
 |||:::
 QY 201 DQAYDDKI 208

RESULT 13

ID W27988 standard; Protein; 113 AA.
 AC W27988;
 DE 26-AUG-1998 (first entry)
 DE Amino acid sequence of elongation factor TU.
 KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
 KW Staphylococcus aureus; regulatory element; bacterial gene expression;
 KW vaccine; staphylococcal infection; food poisoning; scaled skin syndrome;
 KW toxic shock syndrome.
 OS Staphylococcus aureus.
 PN W09730070-A1.
 DR 21-AUG-1997.
 PF 19-FEB-1997; U02318.
 PF 20-FEB-1996; US-011888.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Black Mt, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
 DR WPI; 97-424969/39.

DR N-PSDB; T83947.
PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT to isolate antimicrobial compounds, and in vaccines against S.
PT aureus infection
PS Claim 6; Page 391; 989pp; English.
CC The present sequence represents a Staphylococcus aureus protein, that,
CC based on homology with a Bacillus subtilis protein, is believed
CC to be an elongation factor Tu. The DNA sequence was isolated from a
CC library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
CC sequence can be used in the construction of ribozymes and antisense
CC sequences to control the expression of Staphylococcal genes. The DNA
CC sequence is also useful as a source of regulatory elements for the
CC control of bacterial gene expression. The present protein may be used
CC to produce vaccines to enable a host to produce specific antibodies
CC with antibacterial action. These vaccines and antibodies would protect
CC a host against invasion by S. aureus, and conditions relating to
CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC skin syndrome, and toxic shock syndrome.
SQ Sequence 113 AA;

Query Match 72.7%; Score 40; DB 31; Length 113;
Best Local Similarity 50.0%; Pred. No. 4.54e+02;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 91 daqyeeki 98
| |::||
QY 201 DQAYDDKI 208

RESULT 14
ID W77589 standard; Protein: 146 AA.
AC W77589;
DT 30-OCT-1998 (first entry)
DE Elongation factor Tu.
KW Staphylococcus aureus protein; immune response induction; eye infection;
KW antibody production; T-cell immune response; gastrointestinal infection;
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;
KW central nervous system; kidney infection; urinary tract infection;
KW antimicrobial compound identification; broad spectrum antibiotic;
KW therapy.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT Misc_difference 1..146 /note= "residues designated x are unspecified, and
FT represented as Xaa in the specification"
FT
FT
PN EP-841394-A2.
PD 13-MAY-1998.
PF 24-SEP-1997; 307485.
PR 24-SEP-1996; US-027032.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Black Mt, Burnham MKR, Hodgson JE, Knowles DJC,
PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,
PI Ward JM;
DR WPI: 98-252940/23.
DR N-PSDB: V53387.
PT New nucleic acid sequences from Staphylococcus aureus WCUH29 -
PT useful in vaccines and for treatment of bacterial infections of e.g.
PT respiratory tract and central nervous system
PS Claim 11; Page 269-270; 390pp; English.
CC This sequence represents a Staphylococcus aureus protein, that based on
CC homology with a Bacillus subtilis protein, is an elongation factor Tu
CC (Ef-Tu), and is encoded by a DNA sequence of the invention.
CC The DNA sequences were isolated from Staphylococcus aureus WCUH29
CC (NCIMB 40771). Host cells containing the DNA sequences are used to
CC produce polypeptides or fragments. The proteins are used in the treatment
CC of disease, for inducing an immune response by administering them, to
CC produce antibody and/or T-cell immune response. Antagonists of the
CC proteins are used for the inhibition of bacterial polypeptides.
CC Conditions which may be treated include bacterial infections, especially
CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,
CC urinary tract, skin, bones and joints. The proteins can also be used to
CC identify antimicrobial compounds which are broad spectrum antibiotics,

CC especially useful in the treatment of H. pylori infection.
SQ Sequence 146 AA;

Query Match 72.7%; Score 40; DB 34; Length 146;
Best Local Similarity 50.0%; Pred. No. 4.54e+02;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 124 daqyeeki 131
| |::||
QY 201 DQAYDDKI 208

RESULT 15
ID W24297 standard; Protein: 155 AA.
AC W24297;
DT 14-APR-1998 (first entry)
DE Staphylococcus aureus Gene #2 polypeptide sequence 2.
KW Staphylococcus aureus WCUH 29; antagonist; antibacterial; immunogen;
KW vaccine; disease; protection; isolation.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT Misc_difference 5 /note= "Encoded by GNC"
FT Misc_difference 20 /note= "Encoded by NAA"
FT
FT
PN WO9731114-A2.
PD 28-AUG-1997.
PF 25-FEB-1997; G00524.
PR 26-FEB-1996; GB-004045.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Burnham MKR, Hodgson JE;
DR WPI: 97-435166/40.
DR N-PSDB: T80392.
PT New Staphylococcus aureus polynucleotide and polypeptide(s) - for
PT isolating antagonist of the polypeptide(s) useful as anti-bacterials
PS Claim 11; Page 28; 117pp; English.
CC The present sequence represents a novel polypeptide, which is optionally
CC expressed in NCIMB 40771. The polypeptide, and polynucleotide encoding
CC it, are derived from Staphylococcus aureus. Cells expressing ligands
CC binding the polypeptide can be used to isolated candidate compounds
CC that bind and inhibit the activity of the polypeptide. Such compounds
CC can be used as anti-bacterial compounds. The polypeptide may also be
CC used as an immunogen to vaccinate an animal for protection against
CC Staphylococcus aureus caused disease.
SQ Sequence 155 AA;

Query Match 72.7%; Score 40; DB 27; Length 155;
Best Local Similarity 50.0%; Pred. No. 4.54e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 125 daqygnkl 132
| |::||
QY 201 DQAYDDKI 208

Search completed: Thu Jul 8 18:30:53 1999
Job time : 20 secs.

WORLD

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:30:05 1999; Maspar time 5.05 Seconds
Tabular output not generated. 63.471 Million cell updates/sec

Title: >US-09-041-236-2
Description: (201-208) from US09041236.pap (12 of 45)
Perfect Score: 55
Sequence: 1 DQAYDDKI 8

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.939; Variance 26.599; scale 0.825

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	46	83.6	653	2	T03102 semaphorin homolog A3	1.52e+01
2	45	81.8	80	2	G70187 acyl carrier protein	2.43e+01
3	45	81.8	633	2	S49611 probable serine/threose	2.43e+01
4	45	81.8	1597	2	S68420 citron - mouse	2.43e+01
5	44	80.0	157	2	G59782 conserved hypothetical	3.83e+01
6	44	80.0	645	2	H70173 long-chain-fatty-acid	3.83e+01
7	44	80.0	829	2	S04035 virA protein - Agrob	3.83e+01
8	44	80.0	829	2	B27211 virA protein - Agrob	3.83e+01
9	44	80.0	829	2	S12858 virA protein - Agrob	3.83e+01
10	43	78.2	301	2	A32498 Mx resistance protein	6.01e+01
11	43	78.2	344	2	H71721 hflk protein (hflk)	6.01e+01
12	43	78.2	582	2	S24545 intermediate filament	6.01e+01
13	43	78.2	920	1	PXNCP H+transporing ATPas	6.01e+01
14	43	78.2	922	2	D70066 SNF2 helicase homolog	6.01e+01
15	42	76.4	205	2	E64369 hypothetical protein	9.34e+01
16	42	76.4	248	2	B71862 hypothetical protein	9.34e+01
17	42	76.4	1009	2	S20538 chitin synthase (EC 2	9.34e+01
18	42	76.4	1330	2	S46373 hypothetical protein	9.34e+01
19	42	76.4	2688	2	I49477 alpha-A-crystallin-bi	9.34e+01
20	42	76.4	2717	2	A34203 DNA-binding protein P	9.34e+01
21	41	74.5	144	2	A61158 aminoglycoside N6'-ac	1.44e+02
22	41	74.5	238	2	I64038 hypothetical protein	1.44e+02
23	41	74.5	303	2	S01295 intermediate filament	1.44e+02

24	41	74.5	340	2	P64853 yceG protein precursor	1.44e+02
25	41	74.5	358	2	G64461 biotin synthetase - M	1.44e+02
26	41	74.5	436	2	C69764 4-aminobutyrate amino	1.44e+02
27	41	74.5	450	2	S01294 intermediate filament	1.44e+02
28	41	74.5	472	2	D69755 hypothetical protein	1.44e+02
29	41	74.5	576	2	S12277 intermediate filament	1.44e+02
30	41	74.5	596	2	S51157 intermediate filament	1.44e+02
31	41	74.5	714	2	S76082 hypothetical protein	1.44e+02
32	41	74.5	817	2	S10930 S-receptor kinase (EC	1.44e+02
33	41	74.5	871	1	TNBEA1 97K alpha trans-induc	1.44e+02
34	41	74.5	959	2	T03053 ribonucleoside-diphos	1.44e+02
35	41	74.5	980	2	E71606 hypothetical protein	1.44e+02
36	41	74.5	1396	1	VCBE40 major capsid protein	1.44e+02
37	41	74.5	1408	2	S57049 hypothetical protein	1.44e+02
38	41	74.5	1607	1	MMMSB2 laminin gamma-1 chain	1.44e+02
39	41	74.5	1609	1	MMHUB2 laminin gamma-1 chain	1.44e+02
40	40	72.7	220	3	C45689 reverse transcriptase	2.20e+02
41	40	72.7	251	2	C69630 heptaprenyl diphospha	2.20e+02
42	40	72.7	371	2	S12116 fatty-acyl-CoA reduct	2.20e+02
43	40	72.7	716	1	A35269 translation initiatio	2.20e+02
44	40	72.7	809	2	S67153 probable membrane pro	2.20e+02
45	40	72.7	1777	3	T00490 nonstructural protein	2.20e+02

ALIGNMENTS

RESULT 1
ENTRY T03102 #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
24-Mar-1999
ACCESSIONS T03102
REFERENCE Z14840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-653 #label ENS
#cross-references EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY #length 653 #molecular-weight 73645 #checksum 5501
Query Match 83.6%; Score 46; DB 2; Length 653;
Best Local Similarity 62.5%; Pred. No. 1.52e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 255 NESYDDKI 262
:::|||||
Qy 201 DQAYDDKI 208
RESULT 2
ENTRY G70187 #type complete
TITLE acyl carrier protein homolog - Lyme disease spirochete
ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
21-Aug-1998
ACCESSIONS G70187
REFERENCE A70100
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Winn, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Utterback, L.; Matthey, L.; McDonald, L.; Artiach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
Nature (1997) 390:580-586
Genomic sequence of a Lyme disease spirochaete, Borrelia


```

Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
#cross-references MUID:98044033
#accession G69782
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-157 #label KUN
##cross-references GB:Z99106; GB:Z99107; GB:AL009126; NID:g2632866;
PID:g1182539; PID:g2632873; NID:g2632653;
PID:g1182526; PID:g2632860
##experimental_source strain 168
GENETICS
#gene ydgE
CLASSIFICATION #superfamily Bacillus subtilis conserved hypothetical protein
ydgE
SUMMARY #length 157 #molecular-weight 18374 #checksum 1913
Query Match 80.0%; Score 44; DB 2; Length 157;
Best Local Similarity 50.0%; Pred. No. 3.83e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 44 EQAYENWI 51
:|||||
QY 201 DQAYDDKI 208

RESULT 6
ENTRY H70173 #type complete
TITLE long-chain-fatty-acid CoA ligase homolog - Lyme disease
spirochete
ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease
spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
05-Dec-1998
ACCESSIONS H70173
REFERENCE A70100
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi.
#cross-references MUID:98065943
#accession H70173
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-645 #label KLE
##cross-references GB:AF001161; GB:AE000783; NID:g26689515; PID:g2668524;
TIGR:BB0593
##experimental_source strain B31
CLASSIFICATION #superfamily Synecocystis long-chain-fatty-acid--CoA ligase;
acetate--CoA ligase homology
FEATURE
55-623 #domain acetate--CoA ligase homology #label ACL
SUMMARY #length 645 #molecular-weight 72886 #checksum 1858
Query Match 80.0%; Score 44; DB 2; Length 645;
Best Local Similarity 50.0%; Pred. No. 3.83e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 140 DKSYEKI 147

Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
#cross-references MUID:98044033
#accession G69782
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-157 #label KUN
##cross-references GB:Z99106; GB:Z99107; GB:AL009126; NID:g2632866;
PID:g1182539; PID:g2632873; NID:g2632653;
PID:g1182526; PID:g2632860
##experimental_source strain 168
GENETICS
#gene ydgE
CLASSIFICATION #superfamily Bacillus subtilis conserved hypothetical protein
ydgE
SUMMARY #length 157 #molecular-weight 18374 #checksum 1913
Query Match 80.0%; Score 44; DB 2; Length 157;
Best Local Similarity 50.0%; Pred. No. 3.83e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 44 EQAYENWI 51
:|||||
QY 201 DQAYDDKI 208

RESULT 6
ENTRY H70173 #type complete
TITLE long-chain-fatty-acid CoA ligase homolog - Lyme disease
spirochete
ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease
spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
05-Dec-1998
ACCESSIONS H70173
REFERENCE A70100
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi.
#cross-references MUID:98065943
#accession H70173
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-645 #label KLE
##cross-references GB:AF001161; GB:AE000783; NID:g26689515; PID:g2668524;
TIGR:BB0593
##experimental_source strain B31
CLASSIFICATION #superfamily Synecocystis long-chain-fatty-acid--CoA ligase;
acetate--CoA ligase homology
FEATURE
55-623 #domain acetate--CoA ligase homology #label ACL
SUMMARY #length 645 #molecular-weight 72886 #checksum 1858
Query Match 80.0%; Score 44; DB 2; Length 645;
Best Local Similarity 50.0%; Pred. No. 3.83e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 140 DKSYEKI 147

```

```

QY 201 DQAYDDKI 208
I:||||

RESULT 7
ENTRY S04035 #type complete
TITLE virA protein - Agrobacterium tumefaciens plasmid pTi15955
ORGANISM #formal_name Agrobacterium tumefaciens
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
23-Feb-1997
ACCESSIONS S04035
REFERENCE S04035
#authors Melchers, L.S.; Thompson, D.V.; Idler, K.B.; Neuteboom,
S.T.C.; de Maagd, R.A.; Schilperoort, R.A.; Hooykaas,
P.J.J.
#journal Plant Mol. Biol. (1988) 11:227-237
#title Molecular characterization of the virulence gene virA of the
Agrobacterium tumefaciens octopine Ti plasmid.
#accession S04035
##molecule_type DNA
##residues 1-829 #label MEL
##cross-references EMBL:X12709
GENETICS
#gene virA
#genome plasmid
#keywords autophosphorylation; phosphoprotein; transmembrane protein
SUMMARY #length 829 #molecular-weight 91639 #checksum 7176
Query Match 80.0%; Score 44; DB 2; Length 829;
Best Local Similarity 71.4%; Pred. No. 3.83e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 728 EAYEDKI 734
:|||||
QY 202 QAYDDKI 208

RESULT 8
ENTRY B27211 #type complete
TITLE virA protein - Agrobacterium tumefaciens plasmid pTiA6
ORGANISM #formal_name Agrobacterium tumefaciens
DATE 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change
20-Mar-1998
ACCESSIONS B27211
REFERENCE A91082
#authors Leroux, B.; Yanofsky, M.F.; Winans, S.C.; Ward, J.E.;
Ziegler, S.F.; Nester, E.W.
#journal EMBO J. (1987) 6:849-856
#title Characterization of the virA locus of Agrobacterium
tumefaciens: a transcriptional regulator and host range
determinant.
#cross-references MUID:87246518
#accession B27211
##molecule_type DNA
##residues 1-829 #label LER
##cross-references GB:X05240; NID:g939193; PID:g939194
GENETICS
#gene virA
#genome plasmid
#keywords autophosphorylation; host range; phosphoprotein;
transcription regulation; transmembrane protein
SUMMARY #length 829 #molecular-weight 91797 #checksum 8132
Query Match 80.0%; Score 44; DB 2; Length 829;
Best Local Similarity 71.4%; Pred. No. 3.83e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 728 EAYEDKI 734
:|||||
QY 202 QAYDDKI 208

RESULT 9

```

```
ENTRY          SI2858      #type complete
TITLE          virA protein - Agrobacterium rhizogenes plasmid pRiA4
ORGANISM       #formal_name Agrobacterium rhizogenes
DATE           16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change
              09-Sep-1997
ACCESSIONS     SI2858
REFERENCE      SI2858
#authors      Endoh, H.; Hirayama, T.; Aoyama, T.; Oka, A.
#journal      FEBS Lett. (1990) 271:28-32
#title        Characterization of the virA gene of the agropine-type
              plasmid pRiA4 of Agrobacterium rhizogenes.
#cross-references MUID:91032080
#accession     SI2858
#molecule_type DNA
#residues      1-829 #label END
#cross-references EMBL:X51418; NID:g38991; PID:g38992
GENETICS
#gene          virA
#genome        plasmid
SUMMARY        #length 829 #molecular-weight 91175 #checksum 5655
              80.0%; Score 44; DB 2; Length 829;
              Best Local Similarity 71.4%; Pred. No. 3.83e+01;
              Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 728 EAYDKI 734
:|:|:|
Qy 202 QAYDDKI 208

RESULT 10
ENTRY   A32498      #type fragment
TITLE   Mx resistance protein homolog - perch (fragment)
ORGANISM #formal_name Perca fluviatilis #common_name perch
DATE     25-Sep-1989 #sequence_revision 25-Sep-1989 #text_change
              09-Sep-1997
ACCESSIONS A32498
REFERENCE   A32498
#authors    Staeheli, P.; Yu, Y.X.; Grob, R.; Haller, O.
#journal     Mol. Cell. Biol. (1989) 9:3117-3121
#title       A double-stranded RNA-inducible fish gene homologous to the
              murine influenza virus resistance gene Mx.
#cross-references MUID:89384577
#accession     A32498
#status        preliminary
#molecule_type DNA
#residues      1-301 #label STA
#cross-references GB:M27252; NID:g213517; PID:g213518
SUMMARY        #length 301 #checksum 2504
              78.2%; Score 43; DB 2; Length 301;
              Best Local Similarity 37.5%; Pred. No. 6.01e+01;
              Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 7 NOOYEEKV 14
:|:|:|
Qy 201 DOAYDDKI 208

RESULT 11
ENTRY   H71721      #type complete
TITLE   hfik protein (hflk) RP122 - Rickettsia prowazekii
ORGANISM #formal_name Rickettsia prowazekii
DATE     21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change
              21-Nov-1998
ACCESSIONS H71721
REFERENCE   A71630
#authors    Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.;
              Sichteritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.;
              Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland,
              C.G.
              Nature (1998) 396:133-140
              The genome sequence of Rickettsia prowazekii and the origin
```

```
#accession     H71721      #type complete
#status        preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
#residues      1-344 #label AND
#cross-references GB:AJ235270; GB:AJ235269; NID:g3860572; PID:e1342434;
              PID:g3860690
#experimental_source strain Madrid E
GENETICS
#gene          hflk; RP122
SUMMARY        #length 344 #molecular-weight 39416 #checksum 4078
              78.2%; Score 43; DB 2; Length 344;
              Best Local Similarity 71.4%; Pred. No. 6.01e+01;
              Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 258 QAYNNKI 264
:|:|:|
Qy 202 QAYDDKI 208

RESULT 12
ENTRY   S24545      #type complete
TITLE   intermediate filament protein A, cytosolic - California sea
              hare
ORGANISM #formal_name Aplysia californica #common_name California sea
              hare
DATE     20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
              17-Jul-1998
ACCESSIONS S24545; C60977
REFERENCE   S24545
#authors     Riemer, D.
#submission  submitted to the EMBL Data Library, September 1991
#accession     S24545
#status        preliminary
#molecule_type mRNA
#residues      1-582 #label RIE
#cross-references EMBL:X60297; NID:g5583; PID:g5584
REFERENCE   A60977
#authors     Sweett, J.D.; Kennedy, T.E.; Wager-Smith, K.; Gawinowicz,
              M.A.; Barzilai, A.; Karl, K.A.; Kandel, E.R.
#journal     Electrophoresis (1989) 10:152-157
#title        Development of a database of amino acid sequences for
              proteins identified and isolated on two-dimensional
              polyacrylamide gels.
#cross-references MUID:89276264
#accession     C60977
#molecule_type protein
#residues      'X', 381-399 #label SWE
KEYWORDS      coiled coil; cytosol; intermediate filament
SUMMARY        #length 582 #molecular-weight 65113 #checksum 9790
              78.2%; Score 43; DB 2; Length 582;
              Best Local Similarity 62.5%; Pred. No. 6.01e+01;
              Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 293 NAAYDEKI 300
:|:|:|
Qy 201 DOAYDDKI 208

RESULT 13
ENTRY   PXNCP       #type complete
TITLE   H+-transporting ATPase (EC 3.6.1.35), plasma membrane -
              Neurospora crassa
ALTERNATE_NAMES proton pump
ORGANISM #formal_name Neurospora crassa
DATE     30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
              20-Mar-1998
ACCESSIONS A26497; A25939; A37880; A33846; A50806
REFERENCE   A26497
#authors     Hager, K.M.; Mandala, S.M.; Davenport, J.W.; Speicher, D.W.;
```

```

Benz Jr., E.J.; Slayman, C.W.
Proc. Natl. Acad. Sci. U.S.A. (1986) 83:7693-7697
Amino acid sequence of the plasma membrane ATPase of
Neurospora crassa: deduction from genomic and cDNA
sequences.
#cross-references MUID:87017015
#accession A26497
#molecule_type DNA
#residues 1-920 #label HAG
#cross-references GB:M14085; NID:g168760; PID:g168761
REFERENCE
A25939
#authors Addison, R.
#journal J. Biol. Chem. (1986) 261:14896-14901
#title Primary structure of the Neurospora plasma membrane H+-ATPase
deduced from the gene sequence.
#cross-references MUID:87033715
#accession A25939
#molecule_type DNA
#residues 1-7,'A',9-920 #label ADD
#cross-references GB:J02602
REFERENCE
A37880
#authors Scarborough, G.A.; Hennessey Jr., J.P.
#journal J. Biol. Chem. (1990) 265:16145-16149
#title Identification of the major cytoplasmic regions of the
Neurospora crassa plasma membrane H(+)-ATPase using protein
chemical techniques.
#cross-references MUID:90375470
#accession A37880
#molecule_type protein
#residues 74-81;86-95;190-201;242-252;445-454;475-494;502-508;
549-555;571-582;584-595;616-637;644-659 #label SCA
#note these fragments were released by proteinases from the
cytoplasmic face of proteoliposomes and represent
portions of the cytoplasmic domains of this protein
REFERENCE
A3846
#authors Pardo, J.P.; Slayman, C.W.
#journal J. Biol. Chem. (1989) 264:9373-9379
#title Cysteine 532 and cysteine 545 are the
N-ethylmaleimide-reactive residues of the Neurospora plasma
membrane H(+)-ATPase.
#cross-references MUID:89255436
#accession A3846
#molecule_type protein
#residues 363-375,'X',377-379;520-531,'X',533-537;543-544,'X',
546-548 #label PAR
REFERENCE
A60806
#authors Rao, U.S.; Hennessey Jr., J.P.; Scarborough, G.A.
#journal Anal. Biochem. (1988) 173:251-264
#title Protein chemistry of the Neurospora crassa plasma membrane H
(+)ATPase.
#accession A60806
#molecule_type protein
#residues 272-284,'F',286-291;660-680,'Y',682-698 #label RAO
COMMENT This ATPase transports protons across the plasma membrane to
regulate intracellular pH and ion balance and to generate the
electrochemical gradient that drives nutrient uptake.
COMMENT Topological study of this protein shows certain regions to be
accessible to proteinases on the cytoplasmic face (reference
A37880) and 5 of 8 Cys residues to be accessible to
water-sulphydryl reagents (reference A3846).
GENETICS
#introns 60/3; 129/3; 174/2; 906/1
CLASSIFICATION #superfamily Na+/K+-transporting ATPase alpha chain; ATPase
nucleotide-binding domain homology
KEYWORDS ATP; hydrogen ion transport; hydrolase; phosphoprotein;
transmembrane protein
FEATURE
116-137 #domain transmembrane #status predicted #label TM1\
141-163 #domain transmembrane #status experimental #label TM2\
292-314 #domain transmembrane #status predicted #label TM3\
325-347 #domain transmembrane #status predicted #label TM4\
528-699 #domain ATPase nucleotide-binding domain homology #label
ATN\

```

```

719-743
755-777
827-848
862-884
378
SUMMARY #length 920 #molecular-weight 99886 #checksum 6736
Query Match 78.2%; Score 43; DB 1; Length 920;
Best Local Similarity 62.5%; Pred. No. 6.01e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 495 DQAYKNKV 502
QY 201 DQAYDDKI 208
|||||:
D70066 #type complete
SNP2 helicase homolog ywqA - Bacillus subtilis
#formal_name Bacillus subtilis
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
D70066
A69580
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Broutillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Patro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Presecap, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takamaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
#journal The complete genome sequence of the Gram-positive bacterium
#title Bacillus subtilis.
#cross-references MUID:98044033
#accession D70066
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-922 #label KUN
#cross-references GB:Z99122; GB:AL009136; NID:g2636029; PID:el184534;
PID:g2636153
#experimental_source strain 168
GENETICS

```

```
#gene      ywqa
SUMMARY    #length 922  #molecular-weight 106029  #checksum 1993
Query Match      78.2%; Score 43; DB 2; Length 922;
Best Local Similarity 85.7%; Pred. No. 6.01e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Db 238 QAYSDKI 244
   ||| |||
Qy 202 QAYDDKI 208
```

```
RESULT 15
ENTRY   E64369  #type complete
TITLE   hypothetical protein MJ0557 - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE    13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
        10-Oct-1997
ACCESSIONS E64369
REFERENCE  A64300
#authors   Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
           R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton,
           R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
           Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
           Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
           Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann,
           J.L.; Nguyen, D.; Usterback, T.R.; Kelley, J.M.; Peterson,
           J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
           K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
           H.P.; Fraser, C.W.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal   Science (1996) 273:1058-1073
#title     Complete genome sequence of the methanogenic archaeon,
           Methanococcus jannaschii.
#cross-references MUID:96337999
#accession E64369
#status    preliminary; nucleic acid sequence not shown;
```

```
           translation not shown
#molecule_type DNA
#residues    1-205 #label BUL
#cross-references GB:U67505; GB:L77117; NID:g1591262; PID:g1591263;
           TIGR:MJ0557; PID:g1510634
```

```
GENETICS
#map_position REV493188-492571
SUMMARY    #length 205  #molecular-weight 24082  #checksum 1961
```

```
Query Match      76.4%; Score 42; DB 2; Length 205;
Best Local Similarity 57.1%; Pred. No. 9.34e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 81 HTYDDKV 87
   :|||:
Qy 202 QAYDDKI 208
```

```
Search completed: Thu Jul 8 18:30:15 1999
Job time : 10 secs.
```



```
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 155 PUTATIVE ODORANT-BINDING PROTEIN A10.
SQ SEQUENCE 155 AA; 17828 MW; B96CF633 CRC32;

Query Match 87.3%; Score 48; DB 1; Length 155;
Best Local Similarity 75.0%; Pred. No. 1.55e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 47 EQAYDDKF 54
   :|||||:
QY 201 DQAYDDKI 208

RESULT 2
ID EFTU-NEIGO STANDARD; PRT: 394 AA.
AC P48864;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR TU (EF-TU).
GN TUFA.
OS NEISSERIA GONORRHOAE.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIACEAE; NEISSERIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS11;
RX MEDLINE; 96425873.
RA PORCELLA S.F., BELLAND R.J., JUDD R.C.;
RT "Identification of an EF-Tu protein that is periplasm-associated and
RT processed in Neisseria gonorrhoeae.";
RL MICROBIOLOGY 142:2481-2489(1996).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC Aminoacyl-tRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL: L36380; G790859; ALT_SEQ.
DR PROSITE; PS00301; EFATOR_GTP; 1.
DR PFAM; PF00009; GTP_EFTU; 1.
DR HSP; P02990; LEFU.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 19 26 GTP (BY SIMILARITY).
FT NP_BIND 81 85 GTP (BY SIMILARITY).
FT NP_BIND 136 139 GTP (BY SIMILARITY).
SQ SEQUENCE 394 AA; 43084 MW; A2ED7D6 CRC32;

Query Match 81.8%; Score 45; DB 1; Length 394;
Best Local Similarity 62.5%; Pred. No. 7.80e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 182 DAAYEKI 189
   :|||||:
QY 201 DQAYDDKI 208

RESULT 3
ID PKPA-PHYBL STANDARD; PRT: 613 AA.
AC Q01577;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SERINE/THREONINE PROTEIN KINASE PKPA (EC 2.7.1.1.-).
-----
GN PKPA.
OS PHYCOMYCES BLAKESLEEANUS.
OC EUKARYOTA; FUNGI; ZYGOMYCOTA; ZYGOMYCETES; MUCORALES; MUCORACEAE;
OC PHYCOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 1555;
RX MEDLINE; 96120859.
RA RUIZ-PEREZ V.L., MURILLO F.J., TORRES-MARTINEZ S.;
RT "PKPA, a novel Phycomyces blakesleeanus serine/threonine protein
RT kinase.";
RL CURR. GENET. 28:309-316(1995).
CC -!- FUNCTION: SERINE/THREONINE PROTEIN KINASE THAT PROBABLY
CC PARTICIPATES AS AN INTERMEDIATE IN AN INTRACELLULAR SYSTEM
CC CONTROLLING NUCLEAR PROLIFERATION.
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL: Z46636; G575361; -.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING.
FT DOMAIN 17 269 PROTEIN KINASE.
FT NP_BIND ? ? ATP (BY SIMILARITY).
FT BINDING ? ? ATP (BY SIMILARITY).
FT ACT-SITE 140 140 BY SIMILARITY.
SQ SEQUENCE 613 AA; 69594 MW; 9B2077E3 CRC32;

Query Match 81.8%; Score 45; DB 1; Length 613;
Best Local Similarity 37.5%; Pred. No. 7.80e+00;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 187 EQGYNEKV 194
   :|:|:|:|:
QY 201 DQAYDDKI 208

RESULT 4
ID CTRO-MOUSE STANDARD; PRT: 1597 AA.
AC P49025;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CITRON PROTEIN.
GN CIT.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 96128238.
RA MADAULE P., FURUYASHIKI T., REID T., ISHIZAKI T., WATANABE G.,
RA MORII N., NARUMIYA S.;
RT "A novel partner for the GTP-bound forms of rho and rac.";
RL FEBS LETT. 377:243-248(1995).
CC -!- FUNCTION: PUTATIVE RHO/RAC EFFECTOR THAT BINDS TO THE GTP-BOUND
CC FORMS OF RHO AND RAC1. IT PROBABLY BINDS P21 WITH A TIGHTER
CC SPECIFICITY IN VIVO.
CC -!- SUBUNIT: HOMODIMER (PROBABLE).
CC -!- TISSUE SPECIFICITY: A MAJOR SIGNAL WAS OBSERVED IN TESTIS AND
CC BRAIN, BUT IT WAS ALSO DETECTED IN THYMUS, SPLEEN, KIDNEY, HEART
CC AND LUNG.
```


DE GN WIDE HOST RANGE VIRA PROTEIN (EC 2.7.3.-) (WHR VIRA).
OS AGROBACTERIUM TUMEFACIENS.
OG PLASMID PTI15955.
OC BACTERIA: PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA MELCHERS L.S., THOMPSON D.V., IDLER K.B., NEUTEBOOM S.T.C.,
RT MAAGD R.A., SCHLIPEROORT R.A., HOOYKAAS P.J.J.;
RT "Molecular characterization of the virulence gene of the nopaline plasmid,
RT Agrobacterium tumefaciens octopine Ti plasmid.";
RL PLANT MOL. BIOL. 11:227-237(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90014184.
RA MOREL P., POWELL B.S., ROGOWSKY P.M., KADO C.I.;
RT "Characterization of the virA virulence gene of the nopaline plasmid,
RT pTiC38, of Agrobacterium tumefaciens.";
RL MOL. MICROBIOL. 3:1237-1246(1989).
CC -|- FUNCTION: ACTIVATES VIRG. BY PHOSPHORYLATING IT, IN THE
CC PRESENCE OF ACETOSYRINGONE OR HYDROXYSYRINGONE.
CC -|- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16905; G39189; ALT_SEQ.
DR PIR: S04035; S04035.
DR PFAM: PF00512; signal; 1.
KW SENSORY TRANSDUCTION; TRANSFERASE; KINASE; PHOSPHORYLATION;
KW PLASMID; CROWN GALL TUMOR; TRANSMEMBRANE.
FT TRANSMEM 19 37 POTENTIAL.
FT DOMAIN 451 694 TRANSMITTER DOMAIN (POTENTIAL).
FT MOD_RES 474 474 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 829 AA; 91639 MW; 1C37BD3A CRC32;

Query Match 80.0%; Score 44; DB 1; Length 829;
Best Local Similarity 71.4%; Pred. No. 1.31e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 728 EAYDDKI 734
:|:|:|
Qy 202 QAYDDKI 208

RESULT 8 STANDARD; PRT; 301 AA.
ID MX1_PERFL
AC P20593;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE INTERFERON-INDUCED GTP-BINDING PROTEIN MX (FRAGMENT).
OS PERCA FLUVIATILIS (PERCH).
OS PERCA FLUVIATILIS (PERCH).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;
OC PERCOIDEI; PERCIDAE; PERCA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89384577.
RA STAEHEL P., YU Y.X., GROB R., HALLER O.;
RT "A double-stranded RNA-inducible fish gene homologous to the murine
RT influenza virus resistance gene Mx.";
RL MOL. CELL. BIOL. 9:3117-3121(1989).
CC -|- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M27252; G213518; -.
DR PIR: A32498; A32498.
DR PROSITE: PS00410; DYNAMIN; 1.
DR PFAM: PF00350; dynamin; 1.
KW INTERFERON INDUCTION; GTP-BINDING.
FT NON_TER 1 1
FT NP_BIND 42 49 GTP (POTENTIAL).
FT NP_BIND 143 147 GTP (POTENTIAL).
FT NP_BIND 212 215 GTP (POTENTIAL).
SQ SEQUENCE 301 AA; 33325 MW; 288DADA6 CRC32;

Query Match 78.2%; Score 43; DB 1; Length 301;
Best Local Similarity 37.5%; Pred. No. 2.18e+01;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 7 NQYEEKV 14
:|:|:|
Qy 201 QAYDDKI 208

RESULT 9 STANDARD; PRT; 920 AA.
ID PMAL_NEUCR
AC P07038;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).
GN PMA-1.
OS NEUROSPORA CRASSA.
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC SORDARIALES; SORDARIACEAE; NEUROSPORA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87017015.
RA HAGER K.M., MANDALA S.M., DAVENPORT J.W., SPEICHER D.W.,
RA BENZ E.J. JR., SLAYMAN C.W.;
RT "Amino acid sequence of the plasma membrane ATPase of Neurospora
RT crassa: deduction from genomic and cDNA sequences.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:7693-7697(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87033715.
RA ADDISON R.;
RT "Primary structure of the Neurospora plasma membrane H+-ATPase
RT deduced from the gene sequence. Homology to Na+/K+-, Ca2+-, and
RT K+-ATPase.";
RL J. BIOL. CHEM. 261:14896-14901(1986).
RN [3]
RP PARTIAL SEQUENCE.
RX MEDLINE: 89255436.
RA PARDO J.P., SLAYMAN C.W.;
RT "Cysteine 532 and cysteine 545 are the N-ethylmaleimide-reactive
RT residues of the Neurospora plasma membrane H+-ATPase.";
RL J. BIOL. CHEM. 264:9373-9379(1989).
RN [4]
RP PARTIAL SEQUENCE, AND TOPOLOGY.
RX MEDLINE: 90375470.
RA SCARBOROUGH G.A., HENNESSEY J.P. JR.;
RT "Identification of the major cytoplasmic regions of the Neurospora
RT crassa plasma membrane H(+)-ATPase using protein chemical
RT techniques.";
RL J. BIOL. CHEM. 265:16145-16149(1990).
RN [5]
RP TOPOLOGY.
RX MEDLINE: 91317846.
RA RAO U.S., HENNESSEY J.P. JR., SCARBOROUGH G.A.;

RT Identification of the membrane-embedded regions of the Neurospora
RT crassa plasma membrane H(+)-ATPase";
RL J. BIOL. CHEM. 266:14740-14746(1991).
RN [6]
RN TOPOLOGY.
RX MEDLINE; 92345269.
RX RAO U.S., BAUZON D.D., SCARBOROUGH G.A.;
RA "Cytoplasmic location of amino acids 359-440 of the Neurospora crassa
RT plasma membrane H(+)-ATPase";
RL BIOCHIM. BIOPHYS. ACTA 1108:153-158(1992).
RN [7]
RN TOPOLOGY.
RX MEDLINE; 94148902.
RX LIN A., ADDISON R.;
RA "Topology of the Neurospora plasma membrane H(+)-ATPase. Localization
RT of a transmembrane segment.";
RL J. BIOL. CHEM. 269:3887-3890(1994).
CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING
CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKALINIZATION MAY MEDIATE
CC GROWTH RESPONSES.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O -> ADP + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M14085; G158761; -
CC DR EMBL; J02602; G158765; -
CC DR PIR; A26497; PANC. -
CC DR PFAM; PS00154; ATPASE_E1_E2; 1.
CC DR PFAM; PF00122; E1-E2_ATPase; 1.
CC KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
CC ATP-BINDING.
CC FT DOMAIN 1 115 CYTOPLASMIC.
CC FT TRANSMEM 116 136 1 (POTENTIAL).
CC FT DOMAIN 137 140 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 141 160 2.
CC FT DOMAIN 164 291 CYTOPLASMIC.
CC FT TRANSMEM 292 313 3 (POTENTIAL).
CC FT DOMAIN 314 324 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 325 347 4 (POTENTIAL).
CC FT DOMAIN 348 719 CYTOPLASMIC.
CC FT TRANSMEM 720 738 5 (POTENTIAL).
CC FT DOMAIN 743 754 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 755 774 6 (POTENTIAL).
CC FT DOMAIN 775 826 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 827 847 7.
CC FT DOMAIN 848 859 EXTRACELLULAR.
CC FT TRANSMEM 860 876 8 (POTENTIAL).
CC FT DOMAIN 877 920 CYTOPLASMIC.
CC FT MOD_RES 378 378 PHOSPHORYLATION (PROBABLE).
CC FT BINDING 474 474 ATP (BY SIMILARITY).
CC FT CONFLICT 8 8 G -> A (IN REF. 2).
CC FT CONFLICT 801 801 I -> M (IN REF. 2).
CC SEQUENCE 920 AA; 99886 MW; D3C6B9C4 CRC32;
Query Match 78.28; Score 43; DB 1; Length 920;
Best Local Similarity 62.58; Pred. No. 2:18e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 495 DOAYDKI 502
QY 201 DOAYDDKI 208

RESULT 10
ID TPM2_SCHMA STANDARD; PRT; 284 AA.
AC P42638;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TROPOMYOSIN 2 (TWII).
OS SCHISTOSOMA MANSONI (BLOOD FLUKE).
OC EUKARYOTA; METAZOA; PLATHELMINTHES; TREMATODA; DIGenea; STRIGEIDIDA;
OC SCHISTOSOMATOIDEA; SCHISTOSOMATIDAE; SCHISTOSOMA.
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN-LE;
RC MEDLINE; 93292622.
RA WESTON D.S., KEMP W.M.;
RT "Schistosoma mansoni: comparison of cloned tropomyosin antigens
RL shared between adult parasites and Biomphalaria glabrata";
RL EXP. PARASTOL. 76:358-370(1993).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: ASSOCIATED WITH THE PROTECTIVE TEGUMENT
CC SURROUNDING THE SPOROCYST.
CC -!- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M97555; G161133; -
CC DR PFAM; PF00261; Tropomyosin; 1.
CC DR HSP; P01100; LFOS
CC COILED COIL; REPEAT: MULTIGENE FAMILY.
CC SEQUENCE 284 AA; 32696 MW; E66F6B72 CRC32;
Query Match 76.4%; Score 42; DB 1; Length 284;
Best Local Similarity 37.5%; Pred. No. 3:60e+01;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 218 EAYEENI 225
QY 201 DOAYDDKI 208
RESULT 11
ID CHS2-CANAL STANDARD; PRT; 1009 AA.
AC P30572;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE CHITIN SYNTHASE 2 (EC 2.4.1.16) (CHITIN-UDP ACETYL-GLUCOSAMINYL
DE TRANSFERASE 2).
GN CHS2.
OS CANDIDA ALBICANS (YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC CANDIDACEAE; CANDIDA.
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 92219994.
RX CHEN-WU J.L., ZWICKER J., BOWEN A.R., ROBBINS P.W.;
RT "Expression of chitin synthase genes during yeast and hyphal growth
RL phases of Candida albicans";
RL MOL. MICROBIOL. 6:497-502(1992).
CC -!- FUNCTION: FORMATION AND REPAIR OF THE DISK-SHAPED SEPTUM IN YEAST
CC AND THE CROSS WALLS OF THE HYPHAL PHASE.
CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + [1,4-(N-ACETYL-
CC BETA-D-GLUCOSAMINYL)](N) -> UDP + [1,4-(N-ACETYL-BETA-D-
CC GLUCOSAMINYL)](N+1).
CC -!- SUBCELLULAR LOCATION: PLASMA MEMBRANE-BOUND.

RT "High-resolution solution structure of the double Cys2His2 zinc
RL finger from the human enhancer binding protein MBP-1";
CC BIOCHEMISTRY 31:3907-3917(1992).
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
CC 5'-GGGACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
CC IN T-CELL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
CC ZINC-FINGER IN-BETWEEN.
CC -1- SIMILARITY: 70% BETWEEN THE TWO ZINC-FINGER DOMAINS.
CC -1- SIMILARITY: STRONG, TO HIVP2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X51435; G38018; -
CC PIR; A34203; A34203.
CC DR PFAM; PF00096; zf-C2H2; 5.
CC DR PDB; 3ZNF; 15-JAN-92.
CC DR PDB; 4ZNF; 15-JAN-92.
CC DR PDB; 1BBO; 31-OCT-93.
CC DR MM; 194540; -
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
CC DR PFAM; PF00096; zf-C2H2; 5.
CC DR TRANSFAC; T00497; -
CC KW NUCLEAR PROTEIN; 3D-STRUCTURE.
FT DOMAIN 406 456 ZINC-FINGERS.
FT ZN_FING 406 428 C2H2-TYPE.
FT ZN_FING 434 456 C2H2-TYPE.
FT DOMAIN 803 806 POLY-SER.
FT ZN_FING 958 981 C2HC-TYPE (POTENTIAL).
FT DOMAIN 2087 2139 ZINC-FINGERS.
FT ZN_FING 2087 2109 C2H2-TYPE.
FT ZN_FING 2115 2139 C2H2-TYPE.
FT STRAND 2088 2088
FT TURN 2090 2092
FT STRAND 2095 2095
FT HELIX 2099 2108
FT TURN 2109 2109
FT STRAND 2115 2116
FT STRAND 2123 2124
FT HELIX 2127 2135
SQ SEQUENCE 2717 AA; 297217 MW; 0BDE5C24 CRC32;

Query Match 76.4%; Score 42; DB 1; Length 2717;
Best Local Similarity 71.4%; Pred. No. 3.60e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 864 QSFDDKI 870
Qy 202 QAYDDKI 208
I:::|::|

RESULT 14
ID YG26_HAEIN STANDARD; PRT; 238 AA.
AC P44278;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN H11626.
GN H11626.
OS HAEMOPHILUS INFLUENZAE.

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL SCIENCE 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO B.SUBTILIS YWIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32835; G1574467; -
CC TIGR; H11626; -
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 64 84 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
SQ SEQUENCE 238 AA; 27573 MW; 782A2E4B CRC32;

Query Match 74.5%; Score 41; DB 1; Length 238;
Best Local Similarity 50.0%; Pred. No. 5.86e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 134 DRFDEKI 141
Qy 201 QAYDDKI 208
I:::|::|

RESULT 15
ID YCEG_ECOLI STANDARD; PRT; 340 AA.
AC P28306; P75944;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION.
GN YCEG.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL SCIENCE 277:1453-1474(1997).
RN [2]
RP SEQUENCE OF 1-243 FROM N.A.

RC STRAIN-K12 / W3110;
 RX MEDLINE; 92355506.
 RA GREEN J.M., MERKEL W.K., NICHOLS B.P.;
 RT "Characterization and sequence of Escherichia coli pabc, the gene
 RT encoding aminodeoxychorismate lyase, a pyridoxal phosphate-containing
 RT enzyme.";
 RL J. BACTERIOL. 174:5317-5323(1992).
 CC -!- SIMILARITY: STRONG, TO H.INFLUENZAE H10457.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE000210; G178739; -.
 DR EMBL; M93135; G147061; -.
 DR PIR; B42954; B42954.
 DR PIR; S27567; S27567.
 DR ECGENE; EGI1494; YCEG.
 KW HYPOTHETICAL PROTEIN
 SQ SEQUENCE 340 AA: 38247 MW: 817E2A21 CRC32;

Query Match 74.5%; Score 41; DB 1; Length 340;
 Best Local Similarity 62.5%; Pred. No. 5.86e+01;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 55 EQLYADKI 62
 QY 201 DQAYDDKI 208

Search completed: Thu Jul 8 18:29:18 1999
 Job time : 7 secs.

WQREH (TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:29:36 1999; Maspar time 7.74 Seconds
Tabular output not generated. 56.425 Million cell updates/sec

Title: >US-09-041-236-2
Description: (201-208) from US09041236.pep (12 of 45)
Perfect Score: 55
Sequence: 1 DQAYDDKI 8

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spstreml9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 21.367; Variance 24.249; scale 0.881

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	55	100.0	393	11	SEMAPHORIN L (FRAGMENT)	7.18e-02
2	55	100.0	666	4	SEMAPHORIN L	7.18e-02
3	48	87.3	814	5	T23B12.4 PROTEIN.	3.28e-00
4	46	83.6	653	14	SIMILAR TO GENBANK ACC	9.20e+00
5	46	83.6	827	3	HYPOTHETICAL 95.2 KD P	9.20e+00
6	46	83.6	873	5	ZINC FINGER PROTEIN.	9.20e+00
7	45	81.8	80	2	ACYL CARRIER PROTEIN.	1.52e+01
8	45	81.8	171	2	PROTON-TRANSLLOCATING A	1.52e+01
9	45	81.8	1257	10	ROOT-KNOT NEMATOIDE RES	1.52e+01
10	45	81.8	1641	11	CITRON-K KINASE (FRAGM	1.52e+01
11	45	81.8	2055	11	RHO/RAC-INTERACTING CI	1.52e+01
12	44	80.0	120	5	ARYLPHORIN (FRAGMENT).	2.50e+01
13	44	80.0	157	2	VDGE PROTEIN.	2.50e+01
14	44	80.0	571	2	DEOXYRIBODIPYRIMIDINE	2.50e+01
15	44	80.0	645	2	LONG-CHAIN-FATTY-ACID	2.50e+01
16	44	80.0	817	5	SIMILARITY TO HSP90.	2.50e+01
17	44	80.0	829	2	VIRA HOST (PLASMIID PRI	2.50e+01
18	44	80.0	829	2	WIDE HOST RANGE (WHR)	2.50e+01
19	44	80.0	1132	5	TELOMERASE REVERSE TRA	2.50e+01
20	44	80.0	2910	5	RNA POLYMERASE 1.	2.50e+01

21	43	78.2	221	10	064928	EUKARYOTIC CAP-BINDING	4.08e+01
22	43	78.2	582	5	Q16967	NON-NEURONAL CYTOPLASM	4.08e+01
23	43	78.2	604	5	Q20009	F35B12.1 PROTEIN.	4.08e+01
24	43	78.2	612	5	Q44048	ASPARAGINE-RICH PROTEI	4.08e+01
25	43	78.2	622	2	O85456	HYAE.	4.08e+01
26	43	78.2	922	2	P94593	WYQA PROTEIN (MEMBER O	4.08e+01
27	43	78.2	1099	2	O54377	EXONUCLEASE REXB.	4.08e+01
28	42	76.4	121	5	Q17300	ARYLPHORIN (FRAGMENT).	6.59e+01
29	42	76.4	205	1	Q37977	HYPOTHETICAL PROTEIN M	6.59e+01
30	42	76.4	238	10	O81922	PROLINE-RICH PROTEIN.	6.59e+01
31	42	76.4	322	2	Q56952	YFEA.	6.59e+01
32	42	76.4	818	2	O50510	DNA GYRASE-LIKE PROTEI	6.59e+01
33	42	76.4	833	14	P87603	41KBP FRAGMENT FROM LE	6.59e+01
34	42	76.4	1330	3	Q01910	HYPOTHETICAL PROTEIN T	6.59e+01
35	42	76.4	1902	4	Q14122	DNA-BINDING PROTEIN (M	6.59e+01
36	42	76.4	2708	5	O15791	CHLOROQUINE RESISTANCE	6.59e+01
37	42	76.4	2742	5	O15801	CHLOROQUINE RESISTANCE	6.59e+01
38	42	76.4	2819	5	O15792	STRAIN HB3 CG2 (CG2).	6.59e+01
39	41	74.5	144	2	Q03635	AMINOGLYCOSIDASE-6'-AC	1.06e+02
40	41	74.5	346	5	Q28666	T22C1.6 PROTEIN.	1.06e+02
41	41	74.5	499	5	Q94996	PDDIP.	1.06e+02
42	41	74.5	596	5	Q25421	INTERMEDIATE FILAMENT	1.06e+02
43	41	74.5	714	2	Q55559	HYPOTHETICAL 79.4 KD P	1.06e+02
44	41	74.5	782	5	O61073	SARCOPLASMIC/ENDOPLASM	1.06e+02
45	41	74.5	1070	4	O60772	BK125H2.1 (FRAGMENT).	1.06e+02

ALIGNMENTS

RESULT	1				
ID	O88371	PRELIMINARY;	PRT;	393	AA.
AC	O88371;				
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L (FRAGMENT).				
GN	SEMAL.				
OS	MUS MUSCULUS (MOUSE).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;				
OC	SCIUTOGNATHI; MURIDAE; MURINAE; MUS.				
[1]					
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98389619.				
RA	LANGHE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;				
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA				
RT	viruses."				
RL	GENOMICS 51:340-350(1998).				
DR	EMBL; AF030699; G3523117; --				
FT	NON_TER 393 393				
SQ	SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;				

Query Match 100.0%; Score 55; DB 11; Length 393;
Best Local Similarity 100.0%; Pred. No. 7.18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 231 DQAYDDKI 238
| | | | |
Qy 201 DQAYDDKI 208

RESULT	2				
ID	O75326	PRELIMINARY;	PRT;	666	AA.
AC	O75326;				
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L.				
GN	SEMAL.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;				
OC	CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				

```
RX MEDLINE: 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses.";
RL GENOMICS 51:340-350(1998).
DR EMBL: AF030898; G3523115; -.
SQ SEQUENCE 566 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 55; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 7.19e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 233 DQAYDDKI 240
   |||||
QY 201 DQAYDDKI 208

RESULT 3
ID O17003 PRELIMINARY; PRT; 814 AA.
AC O17003;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE T23B12.4 PROTEIN.
GN T23B12.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA DAVIDSON S., WOHLDMANN P., GILLAM B.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF022982; G2384914; -.
SQ SEQUENCE 814 AA; 93216 MW; 82657C70 CRC32;

Query Match 87.3%; Score 48; DB 5; Length 814;
Best Local Similarity 75.0%; Pred. No. 3.28e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 716 DQAYDDKI 723
   |||||
QY 201 DQAYDDKI 208

RESULT 4
ID O64906 PRELIMINARY; PRT; 653 AA.
AC O64906;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
```

```
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALCELAHPHINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RX MEDLINE: 97201573.
RA ENSSER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RX MEDLINE: 97404659.
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U18243; G1000717; -.
SQ SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match 83.6%; Score 46; DB 14; Length 653;
Best Local Similarity 62.5%; Pred. No. 9.20e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 255 NESYDDKI 262
   ::::|
QY 201 DQAYDDKI 208

RESULT 5
ID O43051 PRELIMINARY; PRT; 827 AA.
AC O43051;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 95.2 KD PROTEIN.
GN SPBC3F6.04C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA LYNE M., RAJANDREAM M.A., BARRELL B.G., CHURCHER C.M.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AL022019; E1254579; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 827 AA; 95151 MW; DD0DBDDA CRC32;

Query Match 83.6%; Score 46; DB 3; Length 827;
Best Local Similarity 62.5%; Pred. No. 9.20e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 806 DQAYNEKM 813
   |||||
QY 201 DQAYDDKI 208

RESULT 6
ID Q23717 PRELIMINARY; PRT; 873 AA.
AC Q23717;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ZINC FINGER PROTEIN.
```


OS CRYPTOSPORIDIUM PARVUM.
OC EUKARYOTA: ALVEOLATA: APICOMPLEXA: COCCIDIA: EIMERIIDA;
OC CRYPTOSPORIDIIDAE; CRYPTOSPORIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA HALABI I.M., STEELE M.I., KUHLS T.L., NIDA K., MOSIER D.A.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL: U48717; G1399466; -.
DR PROSITE, PS00518; ZINC_FINGER_C3HC4; 1.
DR PFAM; PF00097; zf-C3HC4; 1.
KW ZINC-FINGER.
SQ SEQUENCE 873 AA; 99449 MW; 6C1F2A21 CRC32;

Query Match 83.6%; Score 46; DB 5; Length 873;
Best Local Similarity 50.0%; Pred. No. 9.20e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 779 NOTYEDRI 786
QY 201 DQAYDDKI 208
:|:|:|:

RESULT 7
ID O51647 PRELIMINARY; PRT; 80 AA.
AC O51647;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ACYL CARRIER PROTEIN.
GN BB0704.
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC BACTERIA: SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE: 98065943.
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S.,
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTTERBACK T., WATTHEY L., McDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
RL NATURE 390:580-586(1997).
DR EMBL: AE001170; G2688626; -.
DR TIGR: BB0704; -.

SQ SEQUENCE 80 AA; 9334 MW; 2C22A177 CRC32;

Query Match 81.8%; Score 45; DB 2; Length 80;
Best Local Similarity 62.5%; Pred. No. 1.52e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 50 EEAFDDKI 57
QY 201 DQAYDDKI 208
:|:|:|:

RESULT 8
ID O50138 PRELIMINARY; PRT; 171 AA.
AC O50138;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PROTON-TRANSLLOCATING ATPASE, B SUBUNIT.
GN ATPF.
OS RUMINOCOCCUS ALBUS.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP: CLOSTRIDIACEAE;
OC RUMINOCOCCUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC;
RA UMEMORI J., MIWA T., NAGAMINE T., OGATA K., TAKENAKA A., HINO T.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AB006151; D1024565; -.
SQ SEQUENCE 171 AA; 19427 MW; 9BC21655 CRC32;

Query Match 81.8%; Score 45; DB 2; Length 171;
Best Local Similarity 37.5%; Pred. No. 1.52e+01;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 70 EODYEERI 77
QY 201 DQAYDDKI 208
:|:|:|:

RESULT 9
ID O81137 PRELIMINARY; PRT; 1257 AA.
AC O81137;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ROOT-KNOT NEMATODE RESISTANCE PROTEIN.
GN MT-1.2.
OS LYCOPERSICON ESCULENTUM (TOMATO).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; SOLANUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. VENT CHERRY; TISSUE=UNINDUCED ROOT;
RX MEDLINE: 98374252.
RA ROSSI M., GOGGIN F.L., MILLIGAN S.B., KALOSHIAN I., ULLMAN D.E.,
RA WILLIAMSON V.M.;
RT "The nematode resistance gene Mi of tomato confers resistance against the potato aphid."
RL PROC. NATL. ACAD. SCI. U.S.A. 95:9750-9754(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=CV. VENT CHERRY; TISSUE=UNINDUCED ROOT;
RA MILLIGAN S.B., BODEAU J., YAGHOUBI J., KALOSHIAN I., ZABEL P.,
RA WILLIAMSON V.M.;
RT "The root knot nematode resistance gene Mi from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes."
RL PLANT CELL 10:0-0(1998).
DR EMBL: AF039682; G3449380; -.

SQ SEQUENCE 1257 AA; 144587 MW; C1EE7984 CRC32;

Query Match 81.8%; Score 45; DB 10; Length 1257;
Best Local Similarity 85.7%; Pred. No. 1.52e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 582 DQGYDDK 588
QY 201 DQAYDDK 207
|:|:|:

RESULT 10
ID O88528 PRELIMINARY; PRT; 1641 AA.
AC O88528;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CITRON-K KINASE (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE: 96128238.
RA MADAULE P., FURUYASHIKI T., REID T., ISHIZAKI T., WATANABE G.,

RA MORII N., NARUMIYA S.;
RT "A novel partner for the GTP-bound forms of rho and rac.";
RL FEBS LETT. 377:243-248(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RA MADALE P., EDA M., WATANABE N., FUJISAWA K., MATSUOKA T., BITO H.,
RA ISHIZAKI T., NARUMIYA S.;
RT "Role of Citron Kinase as a target of the small GTPase Rho in
RT cytokinesis";
RL NATURE 0:0-0(1998).
DR EMBL: AF070066: G3360514; -.
FT NON_TER 1
SQ SEQUENCE 1641 AA; 187659 MW; 840D790A CRC32;

Query Match 81.8%; Score 45; DB 11; Length 1641;
Best Local Similarity 50.0%; Pred.No. 1.52e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 345 EHQYEEKI 352
: | : : : :
Qy 201 DQAYDDKI 208

RESULT 11
ID O88938 PRELIMINARY; PRT: 2055 AA.
AC O88938;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RHO/RAC-INTERACTING CITRON KINASE.
GN CRK.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA DI CUNTO F., CALAUTTI E., HSIAO J., ONG L., TOPLEY G., TURCO E.,
RA DOTTO G.P.;
RT "CRK, a novel tissue-specific ser/thr kinase encompassing the
RT Rho-Rac binding protein Citron.";
RL J. BIOL. CHEM. 0:0-0(1998).
DR EMBL: AF086824: G3599509; -.
SQ SEQUENCE 2055 AA; 235481 MW; D9AD1E4E CRC32;

Query Match 81.8%; Score 45; DB 11; Length 2055;
Best Local Similarity 50.0%; Pred.No. 1.52e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 759 EHQYEEKI 766
: | : : : :
Qy 201 DQAYDDKI 208

RESULT 12
ID Q26232 PRELIMINARY; PRT: 120 AA.
AC Q26232;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE ARYLPHORIN (FRAGMENT).
OS RHODINIA FUGAX.
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; LEPIOPTERA; BOMBYCOIDEA; SATURNIIDAE; RHODINIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KANAGAWA PREFECTURE; TISSUP-WHOLE BODY;
RA SHIMADA T., KURIMOTO Y., KOBAYASHI M.;
RT "Phylogenetic relationship of silkmooths inferred from sequence data
RT of the arylphorin gene";
RL MOL. PHYLOGENET. EVOL. 0:0-0(1995).
DR EMBL: D44490: D1008521; -.

DR PFAM: PF00372; hemocyanin; 1.
FT NON_TER 1
SQ SEQUENCE 120 AA; 14794 MW; D2631F9A CRC32;
Query Match 80.0%; Score 44; DB 5; Length 120;
Best Local Similarity 85.7%; Pred.No. 2.50e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 100 QAYDQKI 106
: | : : : :
Qy 202 QAYDDKI 208

RESULT 13
ID P96703 PRELIMINARY; PRT: 157 AA.
AC P96703;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE YDGE PROTEIN.
GN YDGE.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE: 98000887.
RA BELOIN C., AVORA S., EXLEY R., HIRSCHBEIN L., OGASAWARA N.,
RA KASAHARA Y., ALONSO J.C., LE HEGARAT F.;
RT "Characterization of an lrp-like (lrpC) gene from Bacillus
RT subtilis";
RL MOL. GEN. GENET. 256:63-71(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE: 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA GHM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISEPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLADO R.P., MIZUNO M., MOSTL D., NAKAI S., NOBACK M.,
RA NOONE D., O'REILLY M., OGAWA K., OGAWARA A., OUDEGA B., PARK S.H.,
RA PARRO V., POHL T.M., PORTELLE D., ROPOLLIK S., PRESCOTT A.M.,
RA PRESECAN E., PUJIC P., PURNELLE B., RAPPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADATE Y.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDI B.,
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WABUITT R., WEDLER E., WEDLER H., WEITZNEGGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis";
RL NATURE 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE: 98044033.
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;

RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: A3001488; D1020146; -
 DR EMBL: Z99106; E1182526; -
 DR EMBL: Z99107; E1182539; -
 DR PFAM: PF00583; Acetyltransf; 1.
 SQ SEQUENCE 157 AA; 18374 MW; 450A34FF CRC32;

Query Match 80.0%; Score 44; DB 2; Length 157;
 Best Local Similarity 50.0%; Pred. No. 2.50e+01;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 44 EQAYENMI 51
 :|||:::
 QY 201 DQAYDDKI 208

RESULT 14

ID O84898 PRELIMINARY; PRT; 571 AA.
 AC O84898;
 DT 01-NOV-1998 (TREMBREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
 DE DEOXYRIBODIPYRIMIDINE PHOTOLYASE.
 GN UVRC.
 OS MYCOPLASMA BOVIS.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
 OC MYCOPLASMATACEAE; MYCOPLASMA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PG45;
 RX MEDLINE: 98329280.
 RA SUBRAMANIAM S., BERGONIER D., POUMARAT F., CAPAUL S., SCHLATTER Y.,
 RA NICOLET J., FREY J.;
 RT "Species identification of Mycoplasma bovis and Mycoplasma agalactiae
 based on the uvrC genes by PCR."
 RL MOL. CELL. PROBES 12:161-169(1998).
 DR EMBL: AF003959; G3435082; -
 KW LYASE.
 SQ SEQUENCE 571 AA; 66463 MW; 75E16439 CRC32;

Query Match 80.0%; Score 44; DB 2; Length 571;
 Best Local Similarity 62.5%; Pred. No. 2.50e+01;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 280 EQFYEDKI 287
 :|||:
 QY 201 DQAYDDKI 208

RESULT 15

ID O51539 PRELIMINARY; PRT; 645 AA.
 AC O51539;
 DT 01-JUN-1998 (TREMBREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBREL. 06, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
 DE LONG-CHAIN-FATTY-ACID COA LIGASE.
 GN BB0593.
 OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
 OC BACTERIA: SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE: 98065943.
 RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
 RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
 RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
 RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
 RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
 RA UTTERBACK T., WATTHEY L., McDONALD L., ARTIACH P., BOWMAN C.,
 RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
 RA SMITH H.O., VENTER J.C.;
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
 burgdorferi.";

RL NATURE 390:580-586(1997).
 DR EMBL: AE001161; G2688524; -
 DR PROSITE: PS00455; AMP_BINDING; 1.
 DR TIGR: BB0593; -
 KW LIGASE.
 SQ SEQUENCE 645 AA; 72886 MW; 8CD7ADE8 CRC32;

Query Match 80.0%; Score 44; DB 2; Length 645;
 Best Local Similarity 50.0%; Pred. No. 2.50e+01;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 140 DKSYEEKI 147
 :|||:
 QY 201 DQAYDDKI 208

Search completed: Thu Jul 8 18:29:48 1999
 Job time : 12 secs.

THIS PAGE BLANK (USPTO)

WIREH

(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_Pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:32:35 1999; MasPar time 6.20 Seconds
Tabular output not generated. 34.287 Million cell updates/sec

Title: >US-09-041-236-2
Description: (213-222) from US09041236.pep (13 of 45)
Perfect Score: 66

Sequence: 1 REDNPKNPE 10

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21265608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 14.860; Variance 44.445; scale 0.334

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	48	72.7	1073	28	W32063	6.28e+01
2	48	72.7	1073	28	Human ST receptor pro	6.28e+01
3	48	72.7	1075	8	GC-C	6.28e+01
4	45	68.2	29	14	Ib-AMP N-terminal (aa	1.35e+02
5	45	68.2	333	14	Antimicrobial protein	1.35e+02
6	45	68.2	619	5	Human calcium channel	1.35e+02
7	45	68.2	993	5	Human calcium channel	1.35e+02
8	45	68.2	2161	33	Human calcium channel	1.35e+02
9	45	68.2	2161	14	Human neuronal calciu	1.35e+02
10	45	68.2	2161	6	Sequence of the alpha	1.35e+02
11	45	68.2	2161	33	Human calcium channel	1.35e+02
12	45	68.2	2161	14	Human neuronal calciu	1.35e+02
13	44	66.7	376	16	T. thermophilus gamma	1.74e+02
14	44	66.7	1382	27	Human metastasis-asso	1.74e+02
15	44	66.7	1684	3	Bacterial amylase A-1	1.74e+02
16	43	65.2	626	33	Streptococcus pneumon	2.23e+02

17	43	65.2	1400	17	R92705	2.23e+02
18	43	65.2	1400	9	R44514	2.23e+02
19	43	65.2	3910	12	R66462	2.23e+02
20	43	65.2	3910	7	R38470	2.23e+02
21	43	65.2	3969	10	R52971	2.23e+02
22	43	65.2	51	20	W08148	2.23e+02
23	42	63.6	51	20	W08146	2.86e+02
24	42	63.6	51	20	W08149	2.86e+02
25	42	63.6	51	20	W08147	2.86e+02
26	42	63.6	150	24	W13652	2.86e+02
27	42	63.6	150	14	R76570	2.86e+02
28	42	63.6	158	19	W02386	2.86e+02
29	42	63.6	437	36	W77350	2.86e+02
30	42	63.6	530	38	W90018	2.86e+02
31	42	63.6	652	34	W49879	2.86e+02
32	42	63.6	783	37	W70586	2.86e+02
33	42	63.6	1150	25	W22471	2.86e+02
34	42	63.6	1164	25	W22469	2.86e+02
35	42	63.6	1167	25	W22470	2.86e+02
36	41	62.1	58	39	W67890	3.66e+02
37	41	62.1	132	14	R76574	3.66e+02
38	41	62.1	155	14	R76573	3.66e+02
39	41	62.1	276	3	R12578	3.66e+02
40	41	62.1	276	2	R23654	3.66e+02
41	41	62.1	305	8	R39356	3.66e+02
42	41	62.1	367	36	W74960	3.66e+02
43	41	62.1	480	38	W74814	3.66e+02
44	41	62.1	506	27	W37495	3.66e+02
45	41	62.1	2270	14	R71010	3.66e+02

ALIGNMENTS

RESULT 1
ID W32063 standard; Protein; 1073 AA.
AC W32063; 1998 (first entry)
DE Human ST receptor protein.
KW ST receptor; colorectal cancer; metastasis; tumour; vaccine.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Domain 24..454
FT Domain /note= "extracellular domain"
FT Domain /note= "transmembrane domain"
FT Domain /note= "cytoplasmic domain"
FT Domain /note= "cytoplasmic domain"
PN W09742220-A1.
PD 13-NOV-1997.
PF 02-MAY-1997; U07565.
PR 03-MAY-1996; US-017018.
PA (UYJE-) UNIV JEFFERSON THOMAS.
PI Waldman SA;
DR WPI; 97-558907/51.
DR N-PSDB; T89300.
PT Use of the human ST toxin receptor protein - for developing products
PT for treating individuals suffering from or susceptible to metastatic
PT colon cancer
PS Claim 1; Page 25-28; 35pp; English.
CC This protein comprises human ST receptor, a receptor found on
CC colorectal cells, including local and metastasized colorectal
CC cancer cells. A claimed vaccine composition includes a protein
CC comprising at least one epitope of human ST receptor, especially
CC an epitope of the extracellular domain of the ST receptor, or a
CC nucleic acid molecule that encodes the protein. The vaccines can
CC be used for treating individuals suffering from, or susceptible
CC to, metastatic colorectal cancer. Susceptible individuals may
CC include: those suffering from inflammatory bowel disease,
CC particularly those with ulcerative colitis; those with colonic
CC polyps; those with familial adenomatous polyposis, a heritable
CC mutation predisposing patients to develop large numbers of
CC intestinal polyps; those with Peutz-Jeghers syndrome; those with
CC hereditary nonpolyposis coli, a heritable mutation which

CC predisposes people to develop colon carcinoma; and those with
CC Turcot syndrome-colon carcinoma in conjunction with independent
CC tumours of the central nervous system. The vaccines may comprise
CC DNA vaccines, infectious vector-mediated vaccines, killed or
CC inactivated vaccines, subunit vaccines and haptenised subunit
CC vaccines. 1073 AA;
SQ Sequence 1073 AA;

Query Match 72.7%; Score 48; DB 28; Length 1073;
Best Local Similarity 40.0%; Pred. No. 6.28e+01;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 728 weedpekrpd 737
:|::|:|:
QY 213 REDNPKNPE 222

RESULT 2
ID W37371 standard; Protein; 1073 AA.
AC W37371;
DT 21-MAY-1998 (first entry)
DE Human ST receptor protein.
KW ST receptor; heat stable toxin receptor; colorectal cancer; tumour;
KW metastasis; diagnosis; human.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Domain 24..430
FT /note= "extracellular domain"
FT Region 1031..1093
FT /note= "C-terminal region"
FT W09742506-A1.
FN 13-NOV-1997.
PD 02-MAY-1997; U07467.
PF 03-MAY-1996; US-016564.
PR (UYJE-) UNIV JEFFERSON THOMAS.
PA Carriethers SL, Waldman SA;
PI WPI: 98-008454/01.
DR N-PSDB; T97229.
DR Determining whether an individual has metastasised colorectal cancer
PT cells and origin of tumour cells - by detecting presence of
PT heat-stable toxin receptor on cells in a sample
PS Example 1: Page 46-49; 62pp; English.
CC This protein comprises human heat-stable toxin (ST) receptor
CC protein, a highly specific marker for metastatic colorectal cancer.
CC New in vitro methods for determining whether or not (i) an
CC individual has metastasised colorectal cancer cells, or (ii) a
CC tumour cell is a colorectal cancer cell, comprise the steps of
CC examining a sample of extraintestinal tissue and/or body fluids or
CC tumour cell from an individual to determine whether ST receptor
CC protein is being expressed by cells in the sample, where expression
CC of ST receptor is determined by: (a) detecting the presence of
CC protein in the sample that binds to antibodies that specifically
CC bind to amino acids 24-430 or 1031-1093 of human ST receptor; or
CC (b) detecting the presence of mRNA in the sample that encodes amino
CC acids 24-430 or 1031-1093 of human ST receptor. Expression may be
CC determined by PCR, where the sample is contacted with primers (see
CC T99462-T99531) that selectively hybridise to sequences that encode
CC amino acids 24-430 of the receptor and which selectively amplify
CC mRNA or cDNA that encodes ST receptor protein (see also T97229).
SQ Sequence 1073 AA;

Query Match 72.7%; Score 48; DB 28; Length 1073;
Best Local Similarity 40.0%; Pred. No. 6.28e+01;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 728 weedpekrpd 737
:|::|:|:
QY 213 REDNPKNPE 222

RESULT 3
ID R38861 standard; Protein; 1075 AA.
AC R38861;

DT 08-FEB-1994 (first entry)
DE GC-C.
KW Guanylyl cyclase; GC-C; heat stable; enterotoxin; rat; small intestine;
KW mucosa; polyA+ RNA; PCR; enterotoxin receptor; bacterial enterotoxin;
KW binding domain; antibody; diarrhoea; ligand; antagonist.
OS Rattus rattus.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /note= "signal peptide"
FT Protein 23..1075
FT /note= "Mature GC-C"
FT modified_site 31..33
FT /note= "N-linked glycosylation site"
FT modified_site 74..76
FT /note= "N-linked glycosylation site"
FT modified_site 78..80
FT /note= "N-linked glycosylation site"
FT modified_site 187..189
FT /note= "N-linked glycosylation site"
FT modified_site 194..196
FT /note= "N-linked glycosylation site"
FT modified_site 306..308
FT /note= "N-linked glycosylation site"
FT modified_site 356..358
FT /note= "N-linked glycosylation site"
FT modified_site 401..403
FT /note= "N-linked glycosylation site"
FT domain 433..453
FT /note= "Transmembrane domain"
FN US5237051-A.
PD 17-AUG-1993.
PE 06-DEC-1990; 623033.
PR 06-DEC-1990; US-623033.
PA (UYVA-) UNIV VANDERBILT.
PI Garbers DL, Schulz S;
DR WPI: 93-272183/34.
DR P-PSDB; R38861.
DR New purified enterotoxin receptor protein - used to develop
PT prods. for treating abnormal conditions caused by bacterially
PT released enterotoxin, partic. diarrhoea
PS Claim 2: Fig 1: 26pp; English.
CC This sequence represents guanylyl cyclase, GC-C, which binds heat
CC stable enterotoxin. The DNA encoding this protein was isolated from
CC rat small intestinal mucosa polyA+ RNA by PCR. This protein is an
CC enterotoxin receptor which may be used as a therapeutic to control
CC intestinal fluid permeation as well as abnormal conditions caused
CC by bacterially released enterotoxin. The binding domain of the
CC protein, or antibodies to the protein, can be used to eliminate
CC diarrhoea. The protein may be used to isolate ligands and to screen
CC for antagonists of toxin binding.
SQ Sequence 1075 AA;

Query Match 72.7%; Score 48; DB 8; Length 1075;
Best Local Similarity 40.0%; Pred. No. 6.28e+01;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 730 weedpekrpd 739
:|::|:|:
QY 213 REDNPKNPE 222

RESULT 4
ID R82948 standard; Peptide; 29 AA.
AC R82948;
DT 30-JAN-1996 (first entry)
DE Ib-AMP N-terminal (aa 26-54) propeptide domain.
KW Antimicrobial protein 1; Ib-AMPI; antifungal; fungicide;
KW antibacterial; phytoncide; disease-resistance; antibiotic;
KW preservative; spacer domain.
OS Impatiens balsamina.
PN W09524486-A1.
PD 14-SEP-1995.
PF 09-MAR-1995; G00509.

PR 11-MAR-1994; GB-004807.
PA (ZENE) ZENECA LTD.
PI Attenborough S, Broekaert WF, Osborn RW, Ray JA;
PI Rees SB, Tailor RH;
DR WPI: 95-328277/42.
PT New antimicrobial proteins from Aralia and Impatiens seeds - useful
PT as fungicides or antibiotics in agricultural or pharmaceutical
PT applications
PS Disclosure; Page 42; 64pp; English.
CC Impatiens balsamina (Ib) antimicrobial protein Ib-AMP (R82942) is
CC a polypeptide including 5 propeptide spacer domains (R89243-47)
CC which lie between Ib-AMP1, Ib-AMP2, Ib-AMP3 and Ib-AMP4 regions,
CC as well as N- and C-terminal propeptide spacer domains (R82948-49)
CC and a signal peptide (R82950). DNA encoding these sequences can
CC be incorporated into plant expression vectors.
SQ Sequence 29 AA;

Query Match 68.2%; Score 45; DB 14; Length 29;
Best Local Similarity 40.0%; Pred. No. 1.35e+02;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 6 keeepakpdp 15
:|:| | |:
Qy 213 REDNPKNPE 222

RESULT 5
ID R82942 standard; Protein; 333 AA.
AC R82942;
DT 30-JAN-1996 (first entry)
DE Antimicrobial protein (Ib-AMP).
KW Antimicrobial protein 1; Ib-AMP1; antifungal; fungicide;
KW antibacterial; phytoncide; disease-resistance; antibiotic;
KW preservative; polypeptide.
OS Impatiens balsamina.

FH Key Location/Qualifiers
FT peptide 1..25
FT domain /label= Sig_peptide
FT domain 26..54
FT domain /label= Propeptide_domain
FT domain 55..74
FT domain /label= Ib-AMP3
FT domain 75..102
FT domain /label= Propeptide_domain
FT domain 103..122
FT domain /label= Ib-AMP1
FT domain 123..148
FT domain /label= Propeptide_domain
FT domain 149..168
FT domain /label= Ib-AMP1
FT domain 169..196
FT domain /label= Propeptide_domain
FT domain 197..216
FT domain /label= Ib-AMP1
FT domain 217..232
FT domain /label= Propeptide_domain
FT domain 233..252
FT domain /label= Ib-AMP2
FT domain 253..278
FT domain /label= Propeptide_domain
FT domain 279..298
FT domain /label= Ib-AMP4
FT domain 299..333
FT domain /label= Propeptide_domain

PN W09524486-A1.
PD 14-SEP-1995.
PF 09-MAR-1995; G00509.
PR 11-MAR-1994; GB-004807.
PA (ZENE) ZENECA LTD.
PI Attenborough S, Broekaert WF, Osborn RW, Ray JA;
PI Rees SB, Tailor RH;
DR WPI: 95-328277/42.
DR N-PSDB; T04958.

PT New antimicrobial proteins from Aralia and Impatiens seeds - useful
PT as fungicides or antibiotics in agricultural or pharmaceutical
PT applications
PS Disclosure; Page 37-38; 64pp; English.
CC An Impatiens balsamina (Ib) cDNA clone (given in T04958)
CC encodes a polypeptide (R82942) that includes 7 propeptide domains
CC that are cleaved during processing to yield antimicrobial
CC proteins Ib-AMP1 (3 copies), Ib-AMP2, Ib-AMP3 and Ib-AMP-4
CC (R82938-41).
SQ Sequence 333 AA;

Query Match 68.2%; Score 45; DB 14; Length 333;
Best Local Similarity 40.0%; Pred. No. 1.35e+02;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 31 keeepakpdp 40
:|:| | |:
Qy 213 REDNPKNPE 222

RESULT 6
ID R27651 standard; Protein; 619 AA.
AC R27651;
DT 03-MAR-1993 (first entry)
DE Human calcium channel 27980/13.
KW Plasmid pCA9.3; Ca-flux assay.
OS Homo sapiens.
PN EP-507170-A.
PD 07-OCT-1992.
PF 23-MAR-1992; 104970.
PR 04-APR-1991; DE-110785.
PA (FARB) BAYER AG.
PI Franz J, Rae P, Unterbeck A, Weingaertner B;
DR WPI: 92-333446/41.
DR N-PSDB; Q29271.
PT Cloned human neuronal calcium channel sub-types - useful in
PT calcium flux assays to screen for neurone-specific calcium
PT channel ligands
PS Claim 2; Page 83-85; 101pp; German.
CC Human neuroblastoma cell line, hippocampus, frontal and temporal
CC cortex and visual cortex cDNA banks were screened with a probe
CC containing carp skeletal muscle Ca-channel cDNA. The cDNA clone
CC pCA9.3 was sequenced. The 5' end of the clone begins directly after
CC domain I (amino acid 337) and the 3' end is at amino acid 922. The
CC neuronal calcium channel protein can be used for screening for Ca
CC channel ligands (agonists or antagonists). See also Q29259-Q29275.
SQ Sequence 619 AA;

Query Match 68.2%; Score 45; DB 5; Length 619;
Best Local Similarity 55.6%; Pred. No. 1.35e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 418 reedekdp 426
:|:| | |:
Qy 213 REDNPKNPE 221

RESULT 7
ID R27650 standard; Protein; 993 AA.
AC R27650;
DT 03-MAR-1993 (first entry)
DE Human calcium channel 27980/12.
KW Plasmid pCA3; Ca-flux assay.
OS Homo sapiens.

FH Key Location/Qualifiers
FT misc_difference 42
FT misc_difference 47
FT misc_difference 47
FT EP-507170-A.
PN 07-OCT-1992.
PD 23-MAR-1992; 104970.
PF 04-APR-1991; DE-110785.

PA (FARB) BAYER AG.
 PI Franz J, Rae P, Unterbeck A, Weingaertner B;
 DR WPI; 92-333446/41.
 DR N-PSDB; Q29270.
 PT Cloned human neuronal calcium channel sub-types - useful in
 PT calcium flux assays to screen for neurone-specific calcium
 PT channel ligands
 PS Claim 2; Page 78-82; 101pp; German.
 CC Human neuroblastoma cell line, hippocampus, frontal and temporal
 CC cortex and visual cortex cDNA banks were screened with a probe
 CC containing carp skeletal muscle Ca-channel cDNA. The cDNA clone
 CC pCA3 was sequenced. The 5'end of the clone lies between
 CC Domains II and III. The clone comprises the complete Domains III
 CC and IV and part of the C-terminus coding sequence. The human
 CC neuronal calcium channel protein can be used for screening for Ca
 CC channel ligands (agonists or antagonists). See also Q29259-Q29275.
 SQ Sequence 993 AA;

Query Match 68.2%; Score 45; DB 5; Length 993;
 Best Local Similarity 55.6%; Pred. No. 1.35e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 reedekdp 82
 ||::||:
 QY 213 REDNPKRNP 221

RESULT 8
 ID W63149 standard; Protein; 2161 AA.
 AC W63149;
 DT 12-OCT-1998 (first entry)
 DE Human calcium channel alpha-1D subunit.
 KW Alpha-1D subunit; human; calcium channel; assay; detection;
 KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis.
 OS Homo sapiens.
 PN US5792846-A.
 PD 11-AUG-1998.
 PF 31-MAY-1995; 455543.
 PR 04-APR-1994; US-223305.
 PR 04-APR-1988; US-176899.
 PR 04-APR-1989; US-603751.
 PR 04-APR-1989; WO-001408.
 PR 20-FEB-1990; US-482384.
 PR 30-NOV-1990; US-620250.
 PR 15-AUG-1991; US-745206.
 PR 31-MAY-1995; US-455543.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams ME;
 DR WPI; 98-456192/39.
 DR N-PSDB; V42697.
 PT DNA encoding human calcium channel alpha 1B sub-unit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome
 PS Disclosure; Columns 271-284; 166pp; English.
 CC The present sequence represents the alpha-1D subunit of a human calcium
 CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
 CC that allow controlled entry of calcium ions into cells. This leads
 CC to depolarisation events required for muscle contraction. The recombinant
 CC subunit, when expressed with nucleic acids encoding the complete calcium
 CC channel, can be used in assays for the detection and characterisation of
 CC compounds that modulate the channel. The DNA encoding the subunits can
 CC be alternatively spliced when transcribed, giving more than one form of
 CC the protein from the same transcript, each having slightly different
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
 CC molecules from the serum of an individual with Lambert Eaton Syndrome
 CC (LES) can be used as a diagnostic for the disease.
 SQ Sequence 2161 AA;

Query Match 68.2%; Score 45; DB 33; Length 2161;
 Best Local Similarity 55.6%; Pred. No. 1.35e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 809 reedekdp 817
 ||::||:
 QY 213 REDNPKNP 221

RESULT 9
 ID R71002 standard; Protein; 2161 AA.
 AC R71002;
 DT 30-NOV-1995 (first entry)
 DE Human neuronal calcium channel subunit alpha 1D including alternative.
 DE exon encoding the IS6 transmembrane domain.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 373..406
 FT /label= encoded by alternative exon
 FT W09504822-A.
 PN 16-FEB-1995.
 PD 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 05-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
 DR WPI; 95-090900/12.
 DR N-PSDB; Q84654.
 PT DNA encoding human calcium channel sub-unit(s) - used for
 PT developing prods. for studying calcium channels, e.g. for
 PT obtaining agonists and antagonists
 PS Disclosure; Page 126-127; 285pp; English.
 CC The alpha 1D subunit cDNA has been isolated using fragments of the
 CC rabbit skeletal muscle calcium channel alpha 1 subunit cDNA as a
 CC probe to screen a cDNA library of human neuroblastoma cell line
 CC IMR32, to obtain clone alpha1.36. This clone was used as a probe to
 CC screen additional IMR32 cell cDNA libraries to obtain overlapping
 CC clones, which were then employed for screening until a sufficient
 CC series of clones to span the length of the nt sequence encoding the
 CC human alpha 1D subunit was obt'd. Full-length clones were then
 CC constructed by ligating partial clones. Q84653 shows the nt sequence
 CC of the cDNA encoding the alpha 1D subunit. The Alpha 1D protein has
 CC a calculated Mr of 245,163. It contains four putative internal
 CC repeated sequence regions which represent 24 putative transmembrane
 CC segments. It mediates DHP-sensitive high-voltage, long-lasting
 CC calcium channel activity. Q84654 shows an alternative exon encoding
 CC the IS6 transmembrane domain. The difference occurs in AAs 373-
 CC 406.
 SQ Sequence 2161 AA;

Query Match 68.2%; Score 45; DB 14; Length 2161;
 Best Local Similarity 55.6%; Pred. No. 1.35e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 809 reedekdp 817
 ||::||:
 QY 213 REDNPKNP 221

RESULT 10
 ID R33345 standard; Protein; 2161 AA.
 AC R33345;
 DT 30-JUN-1993 (first entry)
 DE Sequence of the alpha 1D human calcium channel subunit.
 KW Human calcium channel subunit; diagnosis; agonist; antagonist;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN W09304083-A.
 PD 04-MAR-1993.
 PF 14-AUG-1992; U06903.
 PR 15-AUG-1991; US-745206.
 PR 10-APR-1992; US-868354.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams ME;

DR WPI: 93-093936/11.
DR N-PSDB: Q37811.
PT DNA encoding specific human calcium channel sub-units - used for
PT identifying calcium channel agonists and antagonists and
PT diagnosing Lambert Eaton Syndrome
PS Disclosure: Page 93-101; 150pp; English.
CC The alpha 1D subunit cDNA was isolated using fragments of the
CC rabbit skeletal muscle calcium channel alpha 1 subunit cDNA as a
CC probe to screen a cDNA library of a human neuroblastoma cell line,
CC IMR32, to obtain clone alpha 1.36. This clone was used as a probe to
CC screen additional IMR32 cell cDNA libraries to obtain overlapping
CC clones which were then employed for screening until a sufficient
CC series of clones to span the length of the nucleotide sequence
CC encoding the human alpha 1D subunit were obtained, see Q37811.
CC Q37812 provides the sequence of an alternative exon encoding the IS6
CC transmembrane domain of the alpha 1D subunit. The alpha 1D protein
CC has a calculated Mr of 245,163.
SQ Sequence 2161 AA;

Query Match 68.2%; Score 45; DB 6; Length 2161;
Best Local Similarity 55.6%; Pred. No. 1.35e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 809 reedekdp 817
QY 213 REDNPKNP 221
||:: ||:|

RESULT 11
ID W63137 standard; Protein: 2161 AA.
AC W63137;
DT 12-OCT-1998 (first entry)
DE Human calcium channel alpha-1D subunit.
KW Alpha-1D subunit; human; calcium channel; assay; detection;
KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis.
OS Homo sapiens.
PN US5792846-A.
PD 11-AUG-1998. 455543.
PF 31-MAY-1995; US-223305.
PR 04-APR-1994; US-176899.
PR 04-APR-1988; US-603751.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; WO-U01408.
PR 20-FEB-1990; US-482384.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
DR WPI: 98-456192/39.
DR N-PSDB: V42679.
PT DNA encoding human calcium channel alpha 1B subunit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome
PS Example 1; Columns 261-272; 166pp; English.
CC The present sequence represents the alpha-1D subunit of a human calcium
CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
CC that allow controlled entry of calcium ions into cells. This leads
CC to depolarisation events required for muscle contraction. The recombinant
CC subunit, when expressed with nucleic acids encoding the complete calcium
CC channel, can be used in assays for the detection and characterisation of
CC compounds that modulate the channel. The DNA encoding the subunits can
CC be alternatively spliced when transcribed, giving more than one form of
CC the protein from the same transcript, each having slightly different
CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
CC molecules from the serum of an individual with Lambert Eaton Syndrome
CC (LES) can be used as a diagnostic for the disease.
SQ Sequence 2161 AA;

Query Match 68.2%; Score 45; DB 33; Length 2161;
Best Local Similarity 55.6%; Pred. No. 1.35e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Dr 809 reedekdp 817
QY 213 REDNPKNP 221
||:: ||:|

RESULT 13
ID R86800 standard; Protein: 376 AA.
AC R86800;
DT 19-JUN-1996 (first entry)
DE T. thermophilus gamma-glutamyl kinase.
KW Proline; biosynthetic gene system; gamma-glutamyl; kinase;
KW phosphate reductase; fermentation; enzyme; production; improved;
KW Thermus species; bacterium.
OS Thermus thermophilus.
PN J07298883-A.
PD 14-NOV-1995.
PF 02-MAY-1994; 0933302.
PR 02-MAY-1994; JP-0933302.
PA (AJIN) AJINOMOTO KK.
DR WPI: 96-026424/03.
DR N-PSDB: T06978.
PT DNA encoding Thermus thermophilus gamma-glutamyl kinase - used to
PT improve the fermentative and enzymatic prodn. of proline in Thermus
PT sp. microbe
PS Claim 1; Page 12; 19pp; Japanese.
CC The T. thermophilus proline biosynthetic gene system T06978

Db 809 reedekdp 817
QY 213 REDNPKNP 221
||:: ||:|

RESULT 12
ID R71001 standard; Protein: 2161 AA.
AC R71001;
DT 30-NOV-1995 (first entry)
DE Human neuronal calcium channel subunit alpha 1D.
KW Calcium channel subunit; antagonist; agonist; diagnosis;
KW Lambert Eaton Syndrome.
OS Homo sapiens.
PN WO9504822-A.
PD 16-FEB-1995.
PF 11-AUG-1994; US-105536.
PR 11-AUG-1993; US-149097.
PR 05-NOV-1993; US-149097.
PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
DR WPI: 95-090900/12.
DR N-PSDB: Q84653.
PT DNA encoding human calcium channel sub-unit(s) - used for
PT developing prods. for studying calcium channels, e.g. for
PT obtaining agonists and antagonists
PS Disclosure: Page 116-126; 285pp; English.
CC The alpha 1D subunit cDNA has been isolated using fragments of the
CC rabbit skeletal muscle calcium channel alpha 1 subunit cDNA as a
CC probe to screen a cDNA library of human neuroblastoma cell line
CC IMR32, to obtain clone alpha 1.36. This clone was used as a probe to
CC screen additional IMR32 cell cDNA libraries to obtain overlapping
CC clones, which were then employed for screening until a sufficient
CC series of clones to span the length of the nt sequence encoding the
CC human alpha 1D subunit was obt'd. Full-length clones were then
CC constructed by ligating partial clones. Q84653 shows the nt sequence
CC of the cDNA encoding the alpha 1D subunit. The Alpha 1D protein has
CC a calculated Mr of 245,163. It contains four putative internal
CC repeated sequence regions which represent 24 putative transmembrane
CC segments. It mediates DHP-sensitive high-voltage, long-lasting
CC calcium channel activity.
SQ Sequence 2161 AA;

Query Match 68.2%; Score 45; DB 14; Length 2161;
Best Local Similarity 55.6%; Pred. No. 1.35e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 809 reedekdp 817
QY 213 REDNPKNP 221
||:: ||:|

RESULT 13
ID R86800 standard; Protein: 376 AA.
AC R86800;
DT 19-JUN-1996 (first entry)
DE T. thermophilus gamma-glutamyl kinase.
KW Proline; biosynthetic gene system; gamma-glutamyl; kinase;
KW phosphate reductase; fermentation; enzyme; production; improved;
KW Thermus species; bacterium.
OS Thermus thermophilus.
PN J07298883-A.
PD 14-NOV-1995.
PF 02-MAY-1994; 0933302.
PR 02-MAY-1994; JP-0933302.
PA (AJIN) AJINOMOTO KK.
DR WPI: 96-026424/03.
DR N-PSDB: T06978.
PT DNA encoding Thermus thermophilus gamma-glutamyl kinase - used to
PT improve the fermentative and enzymatic prodn. of proline in Thermus
PT sp. microbe
PS Claim 1; Page 12; 19pp; Japanese.
CC The T. thermophilus proline biosynthetic gene system T06978

```

CC contains three ORFs, which encode gamma-glutamyl kinase (R86800),
CC gamma-glutamyl phosphate reductase (R86801) and the protein R86802,
CC respectively. The gene system can be used to improve the
CC fermentative and enzymatic prodn. of proline in a Thermus spp.
CC bacterium.
SQ Sequence 376 AA;

Query Match 66.7%; Score 44; DB 16; Length 376;
Best Local Similarity 55.6%; Pred. No. 1.74e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 184 ceedpknpq 192
Qy 214 EDNPDKNPE 222
|||||

RESULT 14
ID W31867 standard; Protein: 1382 AA.
AC W31867;
DT 27-APR-1998 (first entry)
DE Human metastasis-associated antigen p150.
KW Antigen; p150; metastasis; tumour; cancer; diagnosis; therapy;
human.
OS Homo sapiens.
PN WO9741221-A1.
PD 06-NOV-1997.
PF 18-APR-1997; E01963.
PR 28-FEB-1997; GB-004161.
PS 29-APR-1996; US-016487.
PA (NOVS ) NOVARTIS AG.
PI Bachmann F;
DR WPI; 97-549725/50.
N-PSDB: T93627.
PT Mouse and human metastasis-associated p150-derived antigens - useful
PT for raising antibodies for therapy and diagnosis of tumours
PS Claim 3; Page 86-91; 11pp; English.
CC This protein comprises a novel human 150 kDa cell surface
CC metastasis-associated antigen, designated p150. It is the
CC homologue of a novel murine p150 metastasis-associated antigen (see
CC W31866). Its amino acid sequence was deduced from an isolated cDNA
CC clone (see T93627). A claimed method for determining the
CC metastatic potential of a tumour cell comprises assessing the level
CC of p150 expression in the cell, a higher level of p150 expression
CC being indicative of a lower state of differentiation and
CC consequently a higher metastatic potential. Immunogenic
CC determinants of p150 and antibodies raised against them are
CC useful in the diagnosis and treatment of metastatic potential in
CC tumours.
SQ Sequence 1382 AA;

Query Match 66.7%; Score 44; DB 27; Length 1382;
Best Local Similarity 70.0%; Pred. No. 1.74e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

Db 1216 reen-dkdpe 1224
Qy 213 REDNPDKNPE 222
|||||

RESULT 15
ID R14948 standard; Protein: 1684 AA.
AC R14948;
DT 25-FEB-1992 (first entry)
DE Bacterial amylase A-180.
KW Maltotriose; G5; starch hydrolysis.
PN EP-459383-A.
PD 04-DEC-1991.
PF 28-MAY-1991; 108669.
PR 31-MAY-1990; DE-017595.
PA (CONE ) CONSORT ELEKTROCHEM IND.
PI Schmid G, Candussio A, Bock A;
DR WPI; 91-355676/49.
N-PSDB: Q13939.

```

```

PT New bacterial amylase, A-180 for malto:pentose prodn. - by
PT hydrolysis of starch, providing high yield and modifiable for
PT secretion from host cells
PS Disclosure: Page 7; 21pp; German.
CC The amino acid sequence is that of bacterial amylase A-180 which is
CC used to prepare maltopentose (G5) in high yields by hydrolysis of
CC starch. It can be modified to ensure its excretion, obviating the need
CC to concentrate and purify the enzyme, i.e. the culture supernatant can
CC be used directly for G5 prodn. G5 yields of over 90% are possible,
CC eliminating the need for further G5 purification. G5 is used in medical
CC assays for diagnosis, as carbohydrate sources in liquid feeding compans.
CC and for converting (by esterification) fatty acids to water-soluble
CC form, suitable for use in stable infusion solns.
SQ Sequence 1684 AA;

Query Match 66.7%; Score 44; DB 3; Length 1684;
Best Local Similarity 85.7%; Pred. No. 1.74e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 361 rennpdk 367
Qy 213 REDNPDK 219
|||||

Search completed: Thu Jul 8 18:32:56 1999
Job time : 21 secs.

```

WIREH

(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:32:08 1999; Maspar time 4.66 Seconds
Tabular output not generated. 86.067 Million cell updates/sec

Title: >US-09-041-236-2
Description: (213-222) from US09041236.pap (13 of 45)
Perfect Score: 56
Sequence: 1 REDNPKNPE 10

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.125; Variance 27.205; scale 0.777

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	49	74.2	1372	1	genome polyprotein -	3.33e+00
2	48	72.7	1073	2	heat-stable enterotox	5.34e+00
3	48	72.7	1075	1	heat-stable enterotox	5.34e+00
4	47	71.2	396	2	Repa protein - Bacill	8.50e+00
5	46	69.7	263	2	APR1 antigen - human	1.35e+01
6	46	69.7	755	1	DNA ligase (ATP) (EC	1.35e+01
7	45	68.2	432	2	dnaj protein homolog	2.12e+01
8	45	68.2	1085	2	cobN protein homolog	2.12e+01
9	45	68.2	1517	2	glutamate synthase (N	2.12e+01
10	45	68.2	2161	2	calcium channel alpha	2.12e+01
11	45	68.2	2181	2	calcium channel alpha	2.12e+01
12	44	66.7	103	2	nonhistone chromosoma	3.30e+01
13	44	66.7	186	2	NADH dehydrogenase (u	3.30e+01
14	44	66.7	202	2	nonhistone chromosoma	3.30e+01
15	44	66.7	250	2	L-ascorbate peroxidase	3.30e+01
16	44	66.7	250	2	L-ascorbate peroxidase	3.30e+01
17	44	66.7	311	2	H+-transporting ATP s	3.30e+01
18	44	66.7	976	2	processing endoprotei	3.30e+01
19	44	66.7	1684	2	amylase A-180 - alkali	3.30e+01
20	43	65.2	79	2	lipoprotein homolog -	5.12e+01
21	43	65.2	173	2	hypothetical protein	5.12e+01
22	43	65.2	215	2	ribosomal protein S6	5.12e+01
23	43	65.2	323	2	protein kinase (EC 2.	5.12e+01

24	43	65.2	391	2	S48085	CAJ1 protein - yeast	5.12e+01
25	43	65.2	504	2	S28298	hypothetical protein	5.12e+01
26	43	65.2	630	2	G54226	hypothetical protein	5.12e+01
27	43	65.2	632	2	S73824	hypothetical protein	5.12e+01
28	43	65.2	646	2	D70939	hypothetical protein	5.12e+01
29	43	65.2	671	1	OKBOG	protein kinase (EC 2.	5.12e+01
30	43	65.2	894	2	S54786	multifunctional beta-	5.12e+01
31	43	65.2	1276	2	T02711	axosome-associated pr	5.12e+01
32	43	65.2	1451	2	B41185	alpha-2 macroglobulin	5.12e+01
33	43	65.2	3866	2	B48205	All-1 protein +GTE fo	5.12e+01
34	43	65.2	3869	2	A48205	All-1 protein +GTE fo	5.12e+01
35	43	65.2	3910	2	A44264	trithorax homolog HFX	5.12e+01
36	43	65.2	3968	2	A44265	trithorax homolog HFX	5.12e+01
37	42	63.6	223	2	A55847	ci repressor homolog	7.89e+01
38	42	63.6	494	1	VCCVCE	coat protein - carnat	7.89e+01
39	42	63.6	550	2	S51931	glutamate synthase (f	7.89e+01
40	42	63.6	596	2	JN0895	succinate dehydrogena	7.89e+01
41	42	63.6	948	2	S52739	H+-transporting ATPas	7.89e+01
42	42	63.6	1021	2	T00361	hypothetical protein	7.89e+01
43	42	63.6	1167	1	A35066	streptococcal C3a pep	7.89e+01
44	42	63.6	1979	2	JW0059	muprd protein - mouse	7.89e+01
45	42	63.6	2025	2	JC5020	tetratricopeptide rep	7.89e+01

ALIGNMENTS

RESULT 1
ENTRY #type fragment
TITLE genome polyprotein - bean yellow mosaic virus (fragment)
CONTAINS 6K protein; coat protein; cylindrical inclusion protein;
proteinase N1a; RNA-directed RNA polymerase (EC 2.7.7.48)
N1b
ORGANISM #formal_name bean yellow mosaic virus, BYMV
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
05-Jun-1998
ACCESSIONS S18921; S22484; S11208
REFERENCE S18921
#authors Boye, K.; Stummann, B.M.; Henningsen, K.W.
#submission submitted to the EMBL Data Library, November 1991
#description CDNA sequence of the Bean Yellow Mosaic Virus nuclear
inclusion protein genes.
#accession S18921
#molecule_type genomic RNA
#residues 1-1372 #label BOY
#cross-references EMBL:X63358
REFERENCE S22484
#authors Boye, K.; Stummann, B.M.; Henningsen, K.W.
#journal Plant Mol. Biol. (1992) 18:1203-1205
#title CDNA cloning and sequencing of the bean yellow mosaic virus
nuclear inclusion protein genes.
#cross-references MUID:92288317
#accession S22484
#molecule_type genomic RNA
#residues 146-849, 'R' 852-1099 #label BO2
#cross-references EMBL:X63358
REFERENCE S11208
#authors Boye, K.; Jensen, P.E.; Stummann, B.M.; Henningsen, K.W.
#journal Nucleic Acids Res. (1990) 18:4926
#title Nucleotide sequence of cDNA encoding the BYMV coat protein
gene.
#cross-references MUID:90370489
#accession S11208
#molecule_type genomic RNA
#residues 1100-1372 #label BOV
#cross-references EMBL:X53684; NID:g61217; PID:g61218
CLASSIFICATION #superfamily tobacco etch virus genome polyprotein; DEAD/H
box helicase homolog
KEYWORDS coat protein; cylindrical inclusion protein; genome-linked
protein; inclusion protein; nucleotidyltransferase;
nucleus; phosphoprotein; polyprotein; proteinase
FEATURE
1-92 #product cylindrical inclusion protein (fragment)
#status predicted #label CIP\

```

93-145      #product 6K protein #status predicted #label VPT\
146-579      #product proteinase Nfa #status predicted #label NIA\
580-1099      #product RNA-directed RNA polymerase Nib #status
              predicted #label RRP\
1100-1372      #product coat protein #status predicted #label CPR\
208          #binding_site phosphoryl-RNA (Tyr) (covalent) #status
              predicted
SUMMARY      #length 1372 #checksum 2614

Query Match      74.2%; Score 49; DB 2; Length 1372;
Best Local Similarity 77.8%; Pred. No. 3.33e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1119 EENPDKNSE 1127
      I:|||||:
QY 214 EDNPDKNPE 222

RESULT 2
ENTRY      OYHUHX      #type complete
TITLE      heat-stable enterotoxin receptor precursor - human
ALTERNATE_NAMES      guanylate cyclase C, membrane-bound; guanylate
                      cyclase-coupled enterotoxin Sfa receptor
CONTAINS      guanylate cyclase (EC 4.6.1.2) C
ORGANISM      #formal_name Homo sapiens #common_name man
DATE      30-Jun-1992 #sequence_revision 22-Apr-1995 #text_change
24-Jul-1998
ACCESSIONS      A40940; JQ1279; S65754
REFERENCE      A40940
              de Sauvage, F.J.; Camerato, T.R.; Goeddel, D.V.
              J. Biol. Chem. (1991) 266:17912-17918
              #journal
              #title
              Primary structure and functional expression of the human
              receptor for Escherichia coli heat-stable enterotoxin.
              #cross-references MUID:92011512
              #accession A40940
              #molecule_type mRNA
              #residues 1-1073 #label DES
              #cross-references GB:M73489; NID:g338501; PID:g338502
              #experimental_source colonic cell line T84 (ATCC CCL248)
REFERENCE      JQ1279
              Singh, S.; Singh, G.; Heim, J.M.; Gerzer, R.
              Biochem. Biophys. Res. Commun. (1991) 179:1455-1463
              #journal
              #title
              Isolation and expression of a guanylate cyclase-coupled heat
              stable enterotoxin receptor cDNA from a human colonic cell
              line.
              #cross-references MUID:92028888
              #accession JQ1279
              #molecule_type mRNA
              #residues 1-321, 'R', 323-330, 'V', 332-508, 'Y', 510-542, 'T', 544-1073
              #experimental_source colonic cell line T84
              #label SIN
REFERENCE      S65754
              Mann, E.A.; Jump, M.L.; Giannella, R.A.
              Biochim. Biophys. Acta (1996) 1305:7-10
              #journal
              #title
              Cell line-specific transcriptional activation of the promoter
              of the human guanylyl cyclase C/heat-stable enterotoxin
              receptor gene.
              #accession S65754
              #molecule_type DNA
              #residues 1-72 #label MAN
              #cross-references EMBL:U02030; NID:g1184045; PID:g1184046
GENETICS
              #gene
              GDB:GUC2C
              #cross-references GDB:128987; OMIM:601330
              #map_position 12p12-12p12
CLASSIFICATION      #superfamily membrane-bound guanylate cyclase; guanylate
                      cyclase catalytic domain homology; natriuretic
                      peptide-binding domain homology; protein kinase homology
                      ATP; carbon-oxygen lyase; cGMP biosynthesis; glycoprotein;
                      heat-stable protein; hormone receptor; phosphorus-oxygen
                      lyase; receptor; transmembrane protein
KEYWORDS
FEATURE      1-23
              #domain signal sequence #status predicted #label SIG\
              #binding_site carbohydrate (Asn) (covalent) #status
              predicted
              #length 1075 #molecular-weight 123753 #checksum 956
SUMMARY
              Query Match      72.7%; Score 48; DB 1; Length 1075;
              Best Local Similarity 40.0%; Pred. No. 5.34e+00;
              Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 730 WEEDPEKRPD 739
      I:|||||:
QY 213 REDNPDKNPE 222

RESULT 3
ENTRY      OYRTHX      #type complete
TITLE      heat-stable enterotoxin receptor precursor - rat
ALTERNATE_NAMES      guanylate cyclase C, membrane-bound
                      guanylate cyclase (EC 4.6.1.2) C
CONTAINS      #formal_name Rattus norvegicus #common_name Norway rat
ORGANISM      31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change
28-Feb-1997
ACCESSIONS      A36292
REFERENCE      A36292
              Schulz, S.; Green, C.K.; Yuen, P.S.T.; Garbers, D.L.
              Cell (1990) 63:941-948
              #journal
              #title
              Guanylyl cyclase is a heat-stable enterotoxin receptor.
              #cross-references MUID:91077932
              #accession A36292
              #molecule_type mRNA
              #residues 1-1075 #label SCH
              #cross-references GB:M55636
CLASSIFICATION      #superfamily membrane-bound guanylate cyclase; guanylate
                      cyclase catalytic domain homology; natriuretic
                      peptide-binding domain homology; protein kinase homology
                      ATP; carbon-oxygen lyase; cGMP biosynthesis; glycoprotein;
                      heat-stable protein; hormone receptor; phosphorus-oxygen
                      lyase; transmembrane protein
KEYWORDS
FEATURE      1-22
              #domain signal sequence #status predicted #label SIG\
              #product heat-stable enterotoxin receptor #status
              predicted #label MAR\
              23-429      #domain extracellular #status predicted #label EXT\
              118-424      #domain natriuretic peptide-binding domain homology
                      #label NPB\
              430-453      #domain transmembrane #status predicted #label TM\
              454-1075      #domain intracellular #status predicted #label CYT\
              474-754      #domain protein kinase homology #label KIN\
              778-1006      #domain guanylate cyclase catalytic domain homology
                      #label CAT\
              31,74,78,187,194,
              306,356,401
              #binding_site carbohydrate (Asn) (covalent) #status
              predicted
              #length 1075 #molecular-weight 123753 #checksum 956
SUMMARY
              Query Match      72.7%; Score 48; DB 1; Length 1075;
              Best Local Similarity 40.0%; Pred. No. 5.34e+00;
              Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 730 WEEDPEKRPD 739
      I:|||||:
QY 213 REDNPDKNPE 222

```

```
RESULT 4
ENTRY RepA protein - Bacillus subtilis plasmid prAT11
TITLE #formal_name Bacillus subtilis
ORGANISM 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change
DATE 20-Mar-1998

ACCESSIONS A27891
REFERENCE #type complete
#authors Imanaka, T.; Ishikawa, H.; Aiba, S.
#journal Mol. Gen. Genet. (1986) 205:90-96
#title Complete nucleotide sequence of the low copy number plasmid
prAT11 and replication control by the RepA protein in
Bacillus subtilis.
#cross-references MUID:87089094
#accession A27891
#molecule_type DNA
#residues 1-396 #label IWA
#cross-references GB:X04400; NID:g40076; PID:g580915

GENETICS plasmid
#genome #length 396 #molecular-weight 47578 #checksum 7705
SUMMARY

Query Match 71.2%; Score 47; DB 2; Length 396;
Best Local Similarity 55.6%; Pred. No. 8.50e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 157 EEEPEKPE 165
|:::|
QY 214 EDNPDKNPE 222

RESULT 5
ENTRY I54780
TITLE #type complete
ORGANISM APK1 antigen - human
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
02-Jul-1996

ACCESSIONS I54780
REFERENCE #type complete
#authors Chang, K.; Pastan, I.
#journal Int. J. Cancer (1994) 57:90-97
#title Molecular cloning and expression of a cDNA encoding a protein
detected by the K1 antibody from an ovarian carcinoma
(OVCAR-3) cell line.
#cross-references MUID:94200897
#accession I54780
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-263 #label RES
#cross-references GB:S72904; NID:g633925; PID:g633926
SUMMARY #length 263 #molecular-weight 30530 #checksum 4141

Query Match 69.7%; Score 46; DB 2; Length 263;
Best Local Similarity 50.0%; Pred. No. 1.35e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 97 REDDPNKEQ 106
|:::|
QY 213 REDNPDKNPE 222

RESULT 6
ENTRY LOBYPX
TITLE #type complete
ALTERNATE_NAMES DNA ligase (ATP) (EC 6.5.1.1) - yeast (Saccharomyces
cerevisiae)
ORGANISM cell division control protein CDC9; polydeoxyribonucleotide
synthase (ATP); protein D1497; protein YDL164c
#formal_name Saccharomyces cerevisiae
DATE 30-Sep-1991 #sequence_revision 12-Apr-1996 #text_change
24-Apr-1998
ACCESSIONS S61049; S05830; S67716; S31138; S74315
```

```
REFERENCE S61010
#authors Pohl, T.M.
#submission submitted to the EMBL Data Library, November 1995
#accession S61049
#molecule_type DNA
#residues 1-755 #label POH
#cross-references EMBL:267750; NID:g1061256; PID:g1061275
REFERENCE S05830
#authors Barker, D.G.; White, J.H.M.; Johnston, L.H.
#journal Nucleic Acids Res. (1985) 13:8323-8337
#title The nucleotide sequence of the DNA ligase gene (CDC9) from
Saccharomyces cerevisiae: a gene which is cell-cycle
regulated and induced in response to DNA damage.
#cross-references MUID:86093646
#accession S05830
#molecule_type DNA
#residues 1-68,'E',70-185,'V',187-755 #label BAR
#cross-references EMBL:X03246; NID:g3514; PID:g3515
REFERENCE S67708
#authors Pohl, T.M.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67716
#molecule_type DNA
#residues 1-755 #label POW
#cross-references EMBL:274212; NID:g1431258; PID:e253072; PID:g1431259;
MIPS:YDL164c
#experimental_source strain S288C
REFERENCE S31138
#authors Wehner, E.P.; Rao, E.; Brendel, M.
#journal Mol. Gen. Genet. (1993) 237:351-358
#title Molecular structure and genetic regulation of SFA, a gene
responsible for resistance to formaldehyde in Saccharomyces
cerevisiae, and characterization of its protein product.
#cross-references MUID:93247548
#accession S31138
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 'AGTCPTS',507,'MIKTVISAATVMPRAHRKELTG',533,
'KALKGIGRRSPLODCS',550,'RKQPSR',558,'RPLSTAARNGA',
570,'KEMAPNSPTAC',584,'HTHAET',591,
'AHEPEVASPIFPIGVDIGATNRTGAGDAGHDASGE',610-755
#label WEH
#cross-references EMBL:X68020
#note the nucleotide sequence was submitted to the EMBL Data
Library, August 1992
#note this sequence has been revised in reference S74315
REFERENCE S74315
#authors Wehner, E.
#submission submitted to the EMBL Data Library, November 1994
#accession S74315
#molecule_type DNA
#residues 610-670,'E',672-723,'I',725-755 #label WEW
#cross-references EMBL:X68020; NID:g577609; PID:g600039
GENETICS
#gene SGD:CDC9
#cross-references SGD:S0002323; MIPS:YDL164c
#map_position 4L
CLASSIFICATION #superfamily yeast polydeoxyribonucleotide synthase (ATP)
KEYWORDS DNA repair; ligase; nucleus; phosphoprotein
FEATURE
419 #active_site Lys (covalent AMP-binding) #status
predicted\
576 #binding_site ATP (Lys) #status predicted
SUMMARY #length 755 #molecular-weight 84827 #checksum 7303

Query Match 69.7%; Score 46; DB 1; Length 755;
Best Local Similarity 75.0%; Pred. No. 1.35e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 334 REDSPDKD 341
|:::|
QY 213 REDNPDKN 220
```

```

7
RESULT 7
ENTRY S48438 #type complete
TITLE dnaJ protein homolog YIR004w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein Y1B4w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 06-Feb-1998
ACCESSIONS S48438; S58705; S50876
REFERENCE S48432
#authors Badcock, K.; Churcher, C.
#submission submitted to the EMBL Data Library, August 1994
#accession S48438
#molecule_type DNA
#residues 1-432 #label BAD
#cross-references GB:Z47047; EMBL:Z38062; NID:g603997; PID:g763349; MIPS:YIR004w
REFERENCE S58703
#authors Voss, H.
#submission submitted to the EMBL Data Library, June 1994
#accession S58705
#molecule_type DNA
#residues 1-432 #label VOS
#cross-references EMBL:X79743
REFERENCE S50795
#authors Wiemann, S.; Schwager, C.; Zimmermann, J.; Sander, C.; Ansorge, W.
#journal Yeast (1995) 11:61-78
#title Nucleotide sequence and analysis of the centromeric region of yeast chromosome IX.
#cross-references MUID:95282515
#accession S50876
#status preliminary: nucleic acid sequence not shown
#molecule_type DNA
#residues 1-75 #label VOW
#cross-references EMBL:X79743
GENETICS
#gene SGD:DJP1
#cross-references SGD:S0001443; MIPS:YIR004w
#map_position 9R
CLASSIFICATION #superfamily dnaJ amino-terminal homology
FEATURE 6-71 #domain dnaJ amino-terminal homology #label DNJ
SUMMARY #length 432 #molecular-weight 48574 #checksum 4839
Query Match 68.2%; Score 45; DB 2; Length 432;
Best Local Similarity 55.6%; Pred. No. 2.12e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 32 OEHPDKNPN 40
Qy 214 EDNPDKNPE 222
:::|||||:

8
RESULT 8
ENTRY T03531 #type complete
TITLE COON protein homolog - Rhodobacter capsulatus
ORGANISM #formal_name Rhodobacter capsulatus
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS T03531
REFERENCE Z14955
#authors Vitek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
#journal Proc. Natl. Acad. Sci. U.S.A. (1997) 94:9384-9388
#title Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003.
#accession T03531
#status preliminary: translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-1085 #label VLC

```

```

#cross-references EMBL:AF010496; NID:g3128256; PID:g3128332
GENETICS
#map_position 1
SUMMARY #length 1085 #molecular-weight 115789 #checksum 447
Query Match 68.2%; Score 45; DB 2; Length 1085;
Best Local Similarity 55.6%; Pred. No. 2.12e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 836 REETPDNP 844
Qy 213 REDNPDKNP 221
:::|:|:|

9
RESULT 9
ENTRY F65112 #type complete
TITLE glutamate synthase (NADPH) (EC 1.4.1.13) large chain precursor - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 04-Sep-1998
ACCESSIONS F65112; A29617
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession F65112
#status preliminary: nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-1517 #label BLAT
#cross-references GB:AF000400; GB:U00096; NID:g2367203; PID:g1789605; UWGP:B3212
REFERENCE A91585
#authors Oliver, G.; Gosset, G.; Sanchez-Pescador, R.; Lozoya, E.; Ku, L.M.; Flores, N.; Becerril, B.; Valle, F.; Bollivar, F.
#journal Gene (1987) 60:1-11
#title Determination of the nucleotide sequence for the glutamate synthase structural genes of Escherichia coli K-12.
#cross-references MUID:88152492
#contents K12
#accession A29617
#status not compared with conceptual translation
#molecule_type DNA
#residues 1-104, 'LAFR', 110-139, 'N', 141-219, 221, 'VCRRIC', 228, 'VLSGSCGPASGM', 241-258, 281, 'BNRSAIRITVKSTSPVTANG', 283-459, 'K', 461-573, 'T', 575-862, 'H', 864-1285, 'DARRS', 1291-1372, 'A', 1374-1375, 'R', 1377-1517 #label OLI
#cross-references GB:M18747; NID:g146207; PID:g146208
GENETICS
#gene gltB
#map_position 69 min
CLASSIFICATION #superfamily glutamate synthase (NADPH)
KEYWORDS flavoprotein; glutamate biosynthesis; iron-sulfur protein; NADP; oxidoreductase
SUMMARY #length 1517 #molecular-weight 166708 #checksum 6315
Query Match 68.2%; Score 45; DB 2; Length 1517;
Best Local Similarity 50.0%; Pred. No. 2.12e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 340 WQNNPDWDE 349
Qy 213 REDNPDKNPE 222
:::||||:|

10
RESULT 10

```

```
ENTRY      JH0564      #type complete
TITLE      calcium channel alpha-1D chain - human
ORGANISM   Homo sapiens #common_name man
DATE       30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
           13-Jun-1997
ACCESSIONS JH0564
REFERENCE   Williams, M.E.; Feldman, D.H.; McCue, A.F.; Brenner, R.;
#authors   Velicelebi, G.; Ellis, S.B.; Harpold, M.M.
#journal   Neuron (1992) 8:71-84
#title     Structure and functional expression of alpha1, alpha2, and
           beta subunits of a novel human neuronal calcium channel
           subtype.
#cross-references MUID:92110010
#accession JH0564
#molecule_type mRNA
##residues 1-2161 ##label WIL
##cross-references GB:M76558
##experimental_source neuroblastoma, cell line IMR32
COMMENT    This protein is a subunit of the voltage-dependent calcium channel.
KEYWORDS   transmembrane protein
FEATURE    127-145   #domain transmembrane #status predicted #label IS1\
164-183   #domain transmembrane #status predicted #label IS2\
195-214   #domain transmembrane #status predicted #label IS3\
236-254   #domain transmembrane #status predicted #label IS4\
274-293   #domain transmembrane #status predicted #label IS5\
382-406   #domain transmembrane #status predicted #label IS6\
524-542   #domain transmembrane #status predicted #label I11\
558-577   #domain transmembrane #status predicted #label I12\
586-603   #domain transmembrane #status predicted #label I13\
615-633   #domain transmembrane #status predicted #label I14\
653-672   #domain transmembrane #status predicted #label I15\
728-752   #domain transmembrane #status predicted #label I16\
888-905   #domain transmembrane #status predicted #label S11\
922-941   #domain transmembrane #status predicted #label S21\
954-972   #domain transmembrane #status predicted #label S31\
980-998   #domain transmembrane #status predicted #label S41\
1018-1037 #domain transmembrane #status predicted #label S51\
1128-1152 #domain transmembrane #status predicted #label S61\
1206-1224 #domain transmembrane #status predicted #label VS1\
1240-1259 #domain transmembrane #status predicted #label VS2\
1268-1286 #domain transmembrane #status predicted #label VS3\
1315-1333 #domain transmembrane #status predicted #label VS4\
1353-1372 #domain transmembrane #status predicted #label VS5\
1440-1464 #domain transmembrane #status predicted #label VS6\
SUMMARY    #length 2161 #molecular-weight 245173 #checksum 5574

Query Match      68.2%; Score 45; DB 2; Length 2161;
Best Local Similarity 55.6%; Pred. No. 2.12e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 809 REDEDKDP 817
QY 213 REDNPKNP 221
|||||

RESULT 11
ENTRY   A38198      #type complete
TITLE   calcium channel alpha-1 chain, pancreatic - human
ALTERNATE_NAMES beta-cell-type calcium channel alpha-1 chain;
                neuroendocrine-type calcium channel alpha-1 chain
ORGANISM Homo sapiens #common_name man
DATE     31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
           10-Sep-1997
ACCESSIONS A38198
REFERENCE   Seino, S.; Chen, L.; Seino, M.; Blondel, O.; Takeda, J.;
#authors   Johnson, J.H.; Bell, G.I.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1992) 89:584-588
#title     Cloning of the alpha 1 subunit of a voltage-dependent calcium
           channel expressed in pancreatic beta cells.
#cross-references MUID:92115705
```

```
#accession A38198
#molecule_type mRNA
##residues 1-2181 ##label SEI
##cross-references GB:M83566; NID:gl79751; PID:gl79752
##experimental_source pancreatic beta cells
KEYWORDS   membrane protein; voltage-gated ion channel
SUMMARY    #length 2181 #molecular-weight 247564 #checksum 2708

Query Match      68.2%; Score 45; DB 2; Length 2181;
Best Local Similarity 55.6%; Pred. No. 2.12e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 829 REDEDKDP 837
QY 213 REDNPKNP 221
|||||

RESULT 12
ENTRY   S32176      #type complete
TITLE   nonhistone chromosomal protein - Halococcus morrhuae
ORGANISM Halococcus morrhuae
DATE     19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
           05-Dec-1998
ACCESSIONS S32176
REFERENCE   Leffers, H.
#authors   Submitted to the EMBL Data Library, March 1993
#submission S32176
#accession  S32176
#molecule_type DNA
##residues 1-103 ##label LEF
##cross-references EMBL:X72588; NID:g288192; PID:g288194
CLASSIFICATION #superfamily methanogen chromosomal protein
SUMMARY    #length 103 #molecular-weight 11735 #checksum 3667

Query Match      66.7%; Score 44; DB 2; Length 103;
Best Local Similarity 60.0%; Pred. No. 3.30e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 72 REDAPDKPN 81
QY 213 REDNPKNPE 222
|||||

RESULT 13
ENTRY   S28242      #type complete
TITLE   NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain CI-ASH1 -
                bovine
ORGANISM Bos primigenius taurus #common_name cattle
DATE     22-Nov-1993 #sequence_revision 13-Mar-1997 #text_change
           17-Mar-1999
ACCESSIONS S28242
REFERENCE   Walker, J.E.; Arizmendi, J.M.; Dupuis, A.; Fearnley, I.M.;
#authors   Finel, M.; Medd, S.M.; Pilkington, S.J.; Runswick, M.J.;
           Skehel, J.M.
#journal   J. Mol. Biol. (1992) 226:1051-1072
#title     Sequences of 20 subunits of NADH: ubiquinone oxidoreductase
           from bovine heart mitochondria. Application of a novel
           strategy for sequencing proteins using the polymerase chain
           reaction.
#cross-references MUID:92389317
#accession S28242
##status preliminary
#molecule_type mRNA
##residues 1-186 ##label WAL
##cross-references EMBL:X63209; NID:g233; PID:g234
KEYWORDS   NAD; oxidoreductase
SUMMARY    #length 186 #molecular-weight 21652 #checksum 1296

Query Match      66.7%; Score 44; DB 2; Length 186;
Best Local Similarity 50.0%; Pred. No. 3.30e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

Db 170 RGGDPNKEPE 179
| :|:|:|
QY 213 REDNPDKNPE 222

Best Local Similarity 60.0%; Pred. No. 3.30e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 14
ENTRY S22359 #type complete
TITLE nonhistone chromosomal protein HMG-2a - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-1999
ACCESSIONS S22359; S36684
REFERENCE S22359
#authors Oka, T.; Endo, Y.; Ito, M.; Miyamoto, K.I.; Sasakawa, T.; Suzuki, I.; Natori, Y.
#journal Biochim. Biophys. Acta (1992) 1130:224-226
#title Molecular cloning of chick liver HMG 2a cDNA and developmental expression of HMG 2a mRNA.
#cross-references MUID:92223100

#accession S22359
#molecule_type mRNA
##residues 1-202 #label OKA1
##cross-references EMBL:X63463
REFERENCE S36684
#authors Oka, T.; Endo, Y.; Ito, M.; Miyamoto, K.I.; Sasakawa, T.; Suzuki, I.; Natori, Y.
#submission submitted to the EMBL Data Library, December 1991
#accession S36684

#molecule_type mRNA
##residues 1-98, 'A', 100-202 #label OKA2
##cross-references EMBL:X63463; NID:g63493; PID:g63494
CLASSIFICATION #superfamily nonhistone chromosomal protein HMG-2; HMG box homology
KEYWORDS chromosomal protein; DNA binding; nucleus
FEATURE
6-83 #domain HMG box homology #label HMGI\
90-164 #domain HMG box homology #label HMG2
SUMMARY #length 202 #molecular-weight 23057 #checksum 8167

Query Match 66.7%; Score 44; DB 2; Length 202;
Best Local Similarity 60.0%; Pred. No. 3.30e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 24 REHKKKNPE 33
|:|:|
QY 213 REDNPDKNPE 222

RESULT 15
ENTRY JF0232 #type complete
TITLE L-ascorbate peroxidase (EC 1.11.1.11) - garden strawberry
ALTERNATE_NAMES APX
ORGANISM #formal_name Fragaria x ananassa
DATE 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 24-Sep-1998

ACCESSIONS JF0232
REFERENCE JF0232
#authors Kim, I. J.; Chung, W. I.
#journal Biosci. Biotechnol. Biochem. (1998) 62:1358-1363
#title Isolation of genomic DNA containing a cytosolic ascorbate peroxidase gene (ApXSC) from the strawberry (Fragaria x ananassa).

#accession JF0232
#molecule_type DNA
##residues 1-250 #label KIM
##cross-references GB:AF039953; NID:g2754859; PID:g2754860
GENETICS

#gene ApXSC
#introns 40/2 98/3 120/3 137/1 165/3 192/2 226/3 246/2
KEYWORDS oxidoreductase
SUMMARY #length 250 #molecular-weight 27283 #checksum 7918

Query Match 66.7%; Score 44; DB 2; Length 250;

W P E R E H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:31:11 1999; MasPar time 3.45 Seconds

Tabular output not generated. 81.918 Million cell updates/sec

Title: >US-09-041-236-2
Description: (213-222) from US09041236.pap (13 of 45)

Perfect Score: 66
Sequence: 1 REDNPKNPE 10

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 21.782; Variance 24.532; scale 0.888

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	48	72.7	1072	1	HSE_RAT	HEAT-STABLE ENTEROTOXIN 1.84e+00
2	48	72.7	1073	1	HSE_PIG	HEAT-STABLE ENTEROTOXIN 1.84e+00
3	48	72.7	1073	1	HSE_HUMAN	HEAT-STABLE ENTEROTOXIN 1.84e+00
4	48	72.7	1076	1	HSE_CAVPO	HEAT-STABLE ENTEROTOXIN 1.84e+00
5	47	71.2	395	1	REPA_BACSU	REPA PROTEIN 3.08e+00
6	46	69.7	755	1	DNL1_YEAST	DNA LIGASE I (EC 6.5.1 5.11e+00
7	45	68.2	333	1	AMP1_IMBPA	ANITBACTERIAL PEPTIDES 8.44e+00
8	45	68.2	432	1	Y1S4_YEAST	HYPOTHETICAL 48.6 KD P 8.44e+00
9	45	68.2	1517	1	GLTB_ECOLI	GLUTAMATE SYNTHASE (NA 8.44e+00
10	44	66.7	186	1	NHAG_BOVIN	NADH-UBIQUINONE OXIDOR 1.38e+01
11	44	66.7	201	1	HNGA_CHICK	HIGH MOBILITY GROUP PR 1.38e+01
12	44	66.7	249	1	APX1_PEA	L-ASCORBATE PEROXIDASE 1.38e+01
13	44	66.7	310	1	ATPG_SPIPL	ATP SYNTHASE GAMMA CHA 1.38e+01
14	44	66.7	376	1	PROB_THETH	GLUTAMATE 5-KINASE (EC 1.38e+01
15	44	66.7	976	1	XPR6_YARLI	DIBASIC PROCESSING END 1.38e+01
16	43	65.2	173	1	CMME_HAEN	CYTOSOL C-TYPE BLOC 2.24e+01
17	43	65.2	215	1	RS6_MYCPN	30S RIBOSOMAL PROTEIN 2.24e+01
18	43	65.2	226	1	BASP_BOVIN	BASPI PROTEIN 2.24e+01
19	43	65.2	391	1	CAJ1_YEAST	CAJ1 PROTEIN 2.24e+01
20	43	65.2	504	1	YLF3_CABEL	HYPOTHETICAL 57.7 KD P 2.24e+01
21	43	65.2	630	1	Y242_MYGE	HYPOTHETICAL PROTEIN M 2.24e+01
22	43	65.2	632	1	Y242_MYCPN	HYPOTHETICAL PROTEIN M 2.24e+01
23	43	65.2	670	1	KGPA_BOVIN	CGMP-DEPENDENT PROTEIN 2.24e+01

RESULT	ID	HSE_RAT	STANDARD	PRT	1072 AA
AC	P23897				
DT	01-NOV-1991	(REL. 20, CREATED)			
DT	01-FEB-1996	(REL. 33, LAST SEQUENCE UPDATE)			
DT	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)			
DE	HEAT-STABLE ENTEROTOXIN RECEPTOR PRECURSOR (GC-C) (INTESTINAL				
DE	GUANYLATE CYCLASE) (EC 4.6.1.2) (STA RECEPTOR).				
GN	GUCY2C OR GUC2C.				
OS	RATTUS NORVEGICUS (RAT).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;				
OC	RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SPRAGUE-DAWLEY; TISSUE=SMALL INTESTINE;				
RX	MEDLINE: 91077932.				
RA	SCHULZ S, GREEN C.K., YUEN P.S.T., GARBERS D.L.;				
RT	"Guanylyl cyclase is a heat-stable enterotoxin receptor.";				
RL	CELL 63:941-948(1990).				
CC	-!- FUNCTION: RECEPTOR FOR THE E.COLI HEAT-STABLE ENTEROTOXIN (E.COLI				
CC	ENTEROTOXIN MARKEDLY STIMULATES THE ACCUMULATION OF CGMP IN				
CC	MAMMALIAN CELLS EXPRESSING GC-C). ALSO ACTIVATED BY THE ENDOGENOUS				
CC	PEPTIDE GUANYLIN				
CC	-!- CATALYTIC ACTIVITY: GTP -> 3',5'-CYCLIC GMP + PYROPHOSPHATE.				
CC	-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.				
CC	-!- TISSUE SPECIFICITY: INTESTINE AND POSSIBLY AT LOW LEVELS IN				
CC	ADDITIONAL TISSUES.				
CC	-!- SIMILARITY: SOME SIMILARITY WITH CONSERVED REGION OF CATALYTIC				
CC	DOMAIN OF PROTEIN KINASES.				
CC	-!- SIMILARITY: TO OTHER GUANYLATE CYCLASES (ONLY IN THE INTRACELLULAR				
CC	DOMAIN).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	EMBL; M55636; G548244; --				
DR	PIR; A36292; OYRTHX.				
DR	PROSITE; PS00452; GUANYLATE_CYCLASES; 1.				
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.				
DR	PFAM; PF00069; pkinase; 1.				
DR	PFAM; PF00211; guanylate_cyc; 1.				
DR	PFAM; PF01094; ANF_receptor; 1.				

ALIGNMENTS

DR HSSP: Q02846; LAWL.
KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; LYASE; CGMP SYNTHESIS; SIGNAL.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1072 HEAT-STABLE ENTEROTOXIN RECEPTOR.
FT DOMAIN 23 429 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 430 453 POTENTIAL.
FT DOMAIN 430 453 POTENTIAL.
FT DOMAIN 454 1072 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 488 748 PROTEIN KINASE.
FT DOMAIN 749 1003 CATALYTIC.
FT CARBOHYD 31 31 POTENTIAL.
FT CARBOHYD 74 74 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 187 187 POTENTIAL.
FT CARBOHYD 194 194 POTENTIAL.
FT CARBOHYD 306 306 POTENTIAL.
FT CARBOHYD 356 356 POTENTIAL.
FT CARBOHYD 401 401 POTENTIAL.
SQ SEQUENCE 1072 AA; 123466 MW; 1E0742D6 CRC32;

Query Match 72.7%; Score 48; DB 1; Length 1072;
Best Local Similarity 40.0%; Pred. No. 1.84e+00;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 727 WEEDPEKRPD 736
:|::|:|:
QY 213 REDNPKNPE 222

RESULT 2
ID HSER_PIG STANDARD; PRT: 1073 AA.
AC P55204; Q29050;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HEAT-STABLE ENTEROTOXIN RECEPTOR PRECURSOR (GC-C) (INTESTINAL
GUANYLATE CYCLASE) (EC 4.6.1.2) (STA RECEPTOR).
GN GUCY2C OR GUC2C.
OS SUC SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SMALL INTESTINE;
RX MEDLINE: 95058376.
RA WADA A., HIRAYAMA T., KITAO S., FUJISAWA J.-I., HIDAKA Y.,
SHIMONISHI Y.;
RT "Pig intestinal membrane-bound receptor (guanylyl cyclase) for heat-
stable enterotoxin: cDNA cloning, functional expression, and
characterization.";
RL MICROBIOL. IMMUNOL. 38:535-541(1994).
RN [2]
RP REVISIONS TO 238 AND 509.
RA WADA A.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: RECEPTOR FOR THE E. COLI HEAT-STABLE ENTEROTOXIN (E. COLI
ENTEROTOXIN MARKEDLY STIMULATES THE ACCUMULATION OF CGMP IN
MAMMALIAN CELLS EXPRESSING GC-C). ALSO ACTIVATED BY THE ENDOGENOUS
PEPTIDE GUANYLIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: GTP -> 3',5'-CYCLIC GMP + PYROPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: SOME SIMILARITY WITH CONSERVED REGION OF CATALYTIC
DOMAIN OF PROTEIN KINASES.
CC -!- SIMILARITY: TO OTHER GUANYLATE CYCLASES (ONLY IN THE INTRACELLULAR
DOMAIN).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: D17513; G1408558; -.
DR PROSITE: PS00452; GUANYLATE_CYCLASES; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM: PF00069; pkinase; 1.
DR PFAM: PF00211; guanylate_cyc; 1.
DR PFAM: PF01094; ANF_receptor; 1.
DR HSSP: Q02846; LAWL.
KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; LYASE; CGMP SYNTHESIS; SIGNAL.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1073 HEAT-STABLE ENTEROTOXIN RECEPTOR.
FT DOMAIN 24 433 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 434 454 POTENTIAL.
FT DOMAIN 455 1073 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 489 749 PROTEIN KINASE.
FT DOMAIN 750 1007 CATALYTIC.
FT CARBOHYD 32 32 POTENTIAL.
FT CARBOHYD 43 43 POTENTIAL.
FT CARBOHYD 79 79 POTENTIAL.
FT CARBOHYD 195 195 POTENTIAL.
FT CARBOHYD 284 284 POTENTIAL.
FT CARBOHYD 307 307 POTENTIAL.
FT CARBOHYD 357 357 POTENTIAL.
FT CARBOHYD 402 402 POTENTIAL.
SQ SEQUENCE 1073 AA; 123219 MW; 122DD58E CRC32;

Query Match 72.7%; Score 48; DB 1; Length 1073;
Best Local Similarity 40.0%; Pred. No. 1.84e+00;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 728 WEEDPEKRPD 737
:|::|:|:
QY 213 REDNPKNPE 222

RESULT 3
ID HSER_HUMAN STANDARD; PRT: 1073 AA.
AC P25092;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HEAT-STABLE ENTEROTOXIN RECEPTOR PRECURSOR (GC-C) (INTESTINAL
GUANYLATE CYCLASE) (EC 4.6.1.2) (STA RECEPTOR).
GN GUCY2C OR GUC2C OR STAR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92011512.
RA DE SAUVAGE F.J., CAMERATO T.R., GOEDDEL D.V.;
RT "Primary structure and functional expression of the human receptor
for Escherichia coli heat-stable enterotoxin.";
RL J. BIOL. CHEM. 266:17912-17918(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92028888.
RA SINGH S., SINGH G., HEIM J.M., GERZER R.;
RT "Isolation and expression of a guanylate cyclase-coupled heat stable
enterotoxin receptor cDNA from a human colonic cell line.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 179:1455-1463(1991).
CC -!- FUNCTION: RECEPTOR FOR THE E. COLI HEAT-STABLE ENTEROTOXIN (E. COLI
ENTEROTOXIN MARKEDLY STIMULATES THE ACCUMULATION OF CGMP IN
MAMMALIAN CELLS EXPRESSING GC-C). ALSO ACTIVATED BY THE ENDOGENOUS
PEPTIDE GUANYLIN.
CC -!- CATALYTIC ACTIVITY: GTP -> 3',5'-CYCLIC GMP + PYROPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: SOME SIMILARITY WITH CONSERVED REGION OF CATALYTIC
DOMAIN OF PROTEIN KINASES.
CC -!- SIMILARITY: TO OTHER GUANYLATE CYCLASES (ONLY IN THE INTRACELLULAR
DOMAIN).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL; M73489; G338502; -
 CC EMBL; S57551; E49425; -
 CC PIR; A0940; A0940
 CC PIR; JQ1279; JQ1279.
 CC MIN; 601330; -
 CC PROSITE; PS00452; GUANYLATE_CYCLASES; 1.
 CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 CC PFAM; PF00069; pkinase; 1.
 CC PFAM; PF00211; guanylate_cyc; 1.
 CC PFAM; PF01094; ANF_receptor; 1.
 CC HSP; Q02846; IAWL.
 CC RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; LYASE; CGMP SYNTHESIS; SIGNAL.
 CC SIGNAL 1 23
 FT CHAIN 24 1073 HEAT-STABLE ENTEROTOXIN RECEPTOR.
 FT DOMAIN 24 430 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 431 454 POTENTIAL.
 FT DOMAIN 455 1073 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 489 749 PROTEIN KINASE.
 FT DOMAIN 750 1007 CATALYTIC.
 FT CARBOHYD 32 32 POTENTIAL.
 FT CARBOHYD 75 75 POTENTIAL.
 FT CARBOHYD 79 79 POTENTIAL.
 FT CARBOHYD 195 195 POTENTIAL.
 FT CARBOHYD 284 284 POTENTIAL.
 FT CARBOHYD 307 307 POTENTIAL.
 FT CARBOHYD 345 345 POTENTIAL.
 FT CARBOHYD 357 357 POTENTIAL.
 FT CARBOHYD 402 402 POTENTIAL.
 FT CONFLICT 322 322 A -> R (IN REF. 2).
 FT CONFLICT 331 331 L -> V (IN REF. 2).
 FT CONFLICT 509 509 D -> V (IN REF. 2).
 FT CONFLICT 543 543 N -> T (IN REF. 2).
 CC SEQUENCE 1073 AA; 123368 MW; 6C5C50C9 CRC32;

Query Match 72.7%; Score 48; DB 1; Length 1073;
 Best Local Similarity 40.0%; Pred. No. 1.84e+00;
 Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 728 WEEDPEKRPD 737
 :|:|:|:|:
 QY 213 REDNPKNPE 222

RESULT 4
 ID HSER_CAVPO STANDARD; PRT; 1076 AA.
 AC P70106;
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE HEAT-STABLE ENTEROTOXIN RECEPTOR PRECURSOR (GC-C) (INTESTINAL
 DE GUANYLATE CYCLASE) (EC 4.6.1.2) (STA RECEPTOR) (GUANYLYL CYCLASE C).
 GN GUCY2C OR GUC2C.
 OS CAVIA PORCELLUS (GUINEA PIG).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; HYSTRICOGNATHI; CAVIIDE; CAVIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-COLON;
 RA KRUBOFFER M., CETIN Y., KAEMPF U., FORSMANN W.-G.;
 RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- FUNCTION: RECEPTOR FOR THE E-COLI HEAT-STABLE ENTEROTOXIN (E-COLI
 CC ENTEROTOXIN MARKEDLY STIMULATES THE ACCUMULATION OF CGMP IN
 CC MAMMALIAN CELLS EXPRESSING GC-C). ALSO ACTIVATED BY THE ENDOGENOUS
 CC PEPTIDE GUANYLIN.
 CC -!- CATALYTIC ACTIVITY: GTP = 3',5'-CYCLIC GMP + PYROPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- SIMILARITY: SOME SIMILARITY WITH CONSERVED REGION OF CATALYTIC

CC DOMAIN OF PROTEIN KINASES
 CC -!- SIMILARITY: TO OTHER GUANYLATE CYCLASES (ONLY IN THE INTRACELLULAR
 CC DOMAIN).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; Z74734; E249974; -
 CC PROSITE; PS00452; GUANYLATE_CYCLASES; 1.
 CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 CC PFAM; PF00069; pkinase; 2.
 CC PFAM; PF00211; guanylate_cyc; 1.
 CC PFAM; PF01094; ANF_receptor; 1.
 CC HSP; Q02846; IAWL.
 CC RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; LYASE; CGMP SYNTHESIS; SIGNAL.
 CC SIGNAL 1 23
 FT CHAIN 24 1076 HEAT-STABLE ENTEROTOXIN RECEPTOR.
 FT DOMAIN 24 433 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 434 457 POTENTIAL.
 FT DOMAIN 458 1076 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 492 752 PROTEIN KINASE.
 FT DOMAIN 753 1011 CATALYTIC.
 FT CARBOHYD 35 35 POTENTIAL.
 FT CARBOHYD 82 82 POTENTIAL.
 FT CARBOHYD 191 191 POTENTIAL.
 FT CARBOHYD 198 198 POTENTIAL.
 FT CARBOHYD 287 287 POTENTIAL.
 FT CARBOHYD 306 306 POTENTIAL.
 FT CARBOHYD 310 310 POTENTIAL.
 FT CARBOHYD 348 348 POTENTIAL.
 FT CARBOHYD 360 360 POTENTIAL.
 FT CARBOHYD 405 405 POTENTIAL.
 CC SEQUENCE 1076 AA; 123119 MW; 6BD534BD CRC32;

Query Match 72.7%; Score 48; DB 1; Length 1076;
 Best Local Similarity 40.0%; Pred. No. 1.84e+00;
 Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 731 WEEDPEKRPD 740
 :|:|:|:|:
 QY 213 REDNPKNPE 222

RESULT 5
 ID REPA_BACSU STANDARD; PRT; 396 AA.
 AC P13962;
 DT 01-JAN-1990 (REL. 13, CREATED)
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
 DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
 DE REPA PROTEIN.
 GN REPA.
 OS BACILLUS SUBTILIS.
 OC PLASMID PRATIL.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 OC BACILLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87089094.
 RA IMANAKA T., ISHIKAWA H., AIBA S.;
 RT "Complete nucleotide sequence of the low copy number plasmid PRATIL
 RT and replication control by the RepA protein in *Bacillus subtilis*.";
 RL MOL. GEN. GENET. 205:90-96(1986).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X04400; G580915; -
CC PIR: A27891; A27891.
KW PLASMID; DNA REPLICATION.
SQ SEQUENCE 396 AA; 47578 MW; 248B0C57 CRC32;

Query Match 71.2%; Score 47; DB 1; Length 396;
Best Local Similarity 55.6%; Pred. No. 3.08e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 157 EEPKPKPE 165
:::|::|
QY 214 EDNPKNPE 222

RESULT 6 DNLI-YEAST STANDARD; PRT; 755 AA.
AC P04819; Q12736;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DNA LIGASE I (EC 6.5.1.1) (POLYDEOXYRIBONUCLEOTIDE SYNTHASE (ATP)).
GN CDC9 OR YDL164C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
SQ SEQUENCE FROM N.A.
RX MEDLINE; 86093646.
RA BARKER D.G., WHITE J.H.M., JOHNSTON L.H.;
RT "The nucleotide sequence of the DNA ligase gene (CDC9) from
Saccharomyces cerevisiae: a gene which is cell-cycle regulated and
induced in response to DNA damage";
RL NUCLEIC ACIDS RES. 13:8323-8337(1985).
RN [2]
SQ SEQUENCE FROM N.A.
RC STRAIN=5288C;
RA POHL T.M.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
SQ SEQUENCE OF 610-755 FROM N.A.
RC STRAIN=AH22;
RX MEDLINE; 93247548.
RA WEHNER E.P., RAO E., BRENDEN M.;
RT "Molecular structure and genetic regulation of SFA, a gene
responsible for resistance to formaldehyde in Saccharomyces
cerevisiae, and characterization of its protein product";
RL MOL. GEN. GENET. 237:351-358(1993).
CC -1- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA
RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA.
CC -1- CATALYTIC ACTIVITY: ATP + (DEOXYRIBONUCLEOTIDE)(N) +
(DEOXYRIBONUCLEOTIDE)(M) -- AMP + PYROPHOSPHATE +
(DEOXYRIBONUCLEOTIDE)(N+M).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- CDC9 IS INCLUDED WITHIN THE CATEGORY OF SO-CALLED 'START GENES',
ENCODING PROTEINS WHICH ARE REQUIRED IN EARLY G1, WHEN THE CELL
IS FACED WITH THE OPTION OF INITIATING A FURTHER CELL CYCLE.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X03246; G3515; -
CC EMBL: Z67750; G1061275; -
CC EMBL: Z74212; E253072; -
CC EMBL: X68020; G600039; -

DR PIR: S05830; LQBPX.
DR PIR: S31138; S31138.
DR SGD: L0000249; CDC9.
DR PROSITE: PS00333; DNA_LIGASE_A2; 1.
DR PROSITE: PS00697; DNA_LIGASE_A1; 1.
DR PFAM: PF01068; DNA_ligase; 1.
KW DNA REPAIR; DNA REPLICATION; DNA RECOMBINATION; CELL DIVISION; LIGASE;
KW ATP-BINDING; NUCLEAR PROTEIN.
FT BINDING 419 419 AMP (BY SIMILARITY).
FT CONFLICT 69 69 D -> E (IN REF. 1).
FT CONFLICT 186 186 L -> V (IN REF. 1).
FT CONFLICT 671 671 G -> E (IN REF. 3).
FT CONFLICT 724 724 R -> I (IN REF. 3).
SQ SEQUENCE 755 AA; 84828 MW; AF42850C CRC32;

Query Match 69.7%; Score 46; DB 1; Length 755;
Best Local Similarity 75.0%; Pred. No. 5.11e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 334 REDSPDKD 341
:::|::|
QY 213 REDNPKN 220

RESULT 7
ID AMP_IMPBA STANDARD; PRT; 333 AA.
AC O24006;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ANITBACTERIAL PEPTIDES PRECURSOR (IB-AMPL-4) [CONTAINS: IB-AMPL;
DE IB-AMP2; IB-AMP3; IB-AMP4].
GN IB-AMP.
OS IMPATIENS BALSAMINA.
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHITES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; THEANAEE; ERICALES; BALSAMINACEAE; IMPATIENS.
RN [1]
SQ SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=SEED;
RX MEDLINE; 97450977.
RA TAILOR R.H., AGLAND D.P., ATTENBOROUGH S., CAMMUE B.P.A., EVANS I.J.,
RA OSBORN R.W., RAY J.A., REES S.B., BROEKERT W.F.;
RT "A novel family of small cysteine-rich antimicrobial peptides from
seed of Impatiens balsamina is derived from a single precursor
protein";
RL J. BIOL. CHEM. 272:24480-24487(1997).
RN [2]
SQ STRUCTURE BY NMR OF IB-AMPL.

RX MEDLINE; 98128575.
RA PATEL S.U., OSBORN R.W., REES S.B., THORNTON J.M.;
RT "Structural studies of Impatiens balsamina antimicrobial protein (IB-
AMPL)";
RL BIOCHEMISTRY 37:983-990(1998).
CC -1- FUNCTION: INHIBITORY TO THE GROWTH OF A RANGE OF FUNGI AND
BACTERIA.

CC -1- MASS SPECTROMETRY: MW=3643; METHOD=ELECTROSPRAY; RANGE=103-122.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: Y14369; E332538; -
KW ANTIBIOTIC; FUNCTION; SIGNAL.
FT SIGNAL 1 23 POTENTIAL.
FT PEPTIDE 55 74 IB-AMP3.
FT PEPTIDE 103 122 IB-AMPL.
FT PEPTIDE 149 168 IB-AMPL.
FT PEPTIDE 197 216 IB-AMPL.

```
FT PEPTIDE 233 252 IB-AMP2.
FT PEPTIDE 279 298 IB-AMP4.
FT DISULFID 60 70 BY SIMILARITY.
FT DISULFID 61 74 BY SIMILARITY.
FT DISULFID 108 118
FT DISULFID 109 122
FT DISULFID 154 164
FT DISULFID 155 168
FT DISULFID 202 212
FT DISULFID 203 216
FT DISULFID 238 248
FT DISULFID 239 252
FT DISULFID 284 294 BY SIMILARITY.
FT DISULFID 285 298 BY SIMILARITY.
SQ SEQUENCE 333 AA; 37259 MW; 91DB7822 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 333;
Best Local Similarity 40.0%; Pred. No. 8.44e+00;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 31 KEEPAKPD 40
:|:| | |:|
QY 213 REDNPKNPE 222

RESULT 8
ID Y1S4_YEAST STANDARD; PRT; 432 AA.
AC P40564;
RX MEDLINE; 95282515.
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA CHURCHER C.M., CONNOR R., COPESEY T., DEAR S., DEVLIN K., FRASER A.,
RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA WALSH S.V., WHITEHEAD S.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE; 95282515.
RA VOSS H., TAMAMES J., TEODORU C., VALENCIA A., SENSEN C., WIEMANN S.,
RA SCHWAGER C., ZIMMERMANN J., SANDER C., ANSORGE W.;
RT "Nucleotide sequence and analysis of the centromeric region of yeast
chromosome IX."
RL YEAST 11:61-78(1995).
CC -!- SIMILARITY: CONTAINS A DNAAJ-LIKE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 238062; G557853; -
CC EMBL; 247047; G763349; -
CC EMBL; X79743; -; NOT_ANNOTATED_CDS.
CC PIR; S48438; S48438.
CC PROSITE; PS00636; DNAAJ_1; 1.
CC PFAM; PF00226; DnaJ; 1.
CC HSSP; P08622; 1XBL.

KW HYPOTHETICAL PROTEIN; CHAPERONE.
FT DOMAIN 4 73 DNAAJ-LIKE.
SQ SEQUENCE 432 AA; 48574 MW; E5EDC08E CRC32;

Query Match 68.2%; Score 45; DB 1; Length 432;
Best Local Similarity 55.6%; Pred. No. 8.44e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 32 QEHDPKNPN 40
:|:| | |:|
QY 214 EDNPKNPE 222

RESULT 9
ID GLTB_ECOLI STANDARD; PRT; 1517 AA.
AC P09831;
RX MEDLINE; 88152492.
RA OLIVER G., GOSSET G., SANCHEZ-PESCAOR R., LOZOYA E., KU L.M.,
RA FLORES N., BECERRIL B., VALLE F., BOLIVAR F.;
RT "Determination of the nucleotide sequence for the glutamate synthase
structural genes of Escherichia coli K-12."
RL GENE 60:1-11(1987).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL SCIENCE 277:1453-1474(1997).
[3]
RP SEQUENCE OF 1-304 FROM N.A.
RX MEDLINE; 91217005.
RA VELAZQUEZ L., CAMARENA L., REYES J.L., BASTARRACHEA F.;
RT "Mutations affecting the Shine-Dalgarno sequences of the untranslated
region of the Escherichia coli gitBDF operon."
RL J. BACTERIOL. 173:3261-3264(1991).
[4]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE; 89088858.
RA GOSSET G., MERINO E., RECILLAS F., OLIVER G., BECERRIL B., BOLIVAR F.;
RT "Amino acid sequence analysis of the glutamate synthase enzyme from
Escherichia coli K-12."
RL PROTEIN SEQ. DATA ANAL. 2:9-16(1989).
CC -!- CATALYTIC ACTIVITY: 2 L-GLUTAMATE + NADP(+) -> L-GLUTAMINE +
CC 2-OXOGLUTARATE + NADPH.
CC -!- COFACTOR: IRON-SULFUR; FAD AND FMN FLAVOPROTEIN.
CC -!- PATHWAY: NITROGEN METABOLISM, GLUTAMATE BIOSYNTHESIS.
CC THE CATALYZED REACTION BRINGS TOGETHER THE NITROGEN AND
CC CARBON METABOLISM.
CC -!- SUBUNIT: AGGREGATE OF 4 CATALYTICAL ACTIVE HETERODIMERS,
CC CONSISTING OF A LARGE AND A SMALL SUBUNIT.
CC -!- GLUTAMINE BINDS TO THE LARGE SUBUNIT AND TRANSFERS THE AMIDO GROUP
CC TO 2-OXO-GLUTAMATE THAT APPARENTLY BINDS TO THE SMALL SUBUNIT.
CC -!- SIMILARITY: TO OTHER GLUTAMATE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: M18747; G146208; -
DR EMBL: U18997; G606151; -
DR EMBL: AE000400; G1789605; -
DR EMBL: M68876; G146211; -
DR PIR: A29617; A29617;
DR ECGENE: EG10403; GLTB.
KW OXIDOREDUCTASE; IRON-SULFUR; 3FE-4S; FLAVOPROTEIN; FAD; FMN;
KW GLUTAMATE BIOSYNTHESIS; ZYMOGEN.
FT PROPEP 1 42
FT CHAIN 43 1517 GLUTAMATE SYNTHASE LARGE CHAIN.
FT METAL 1133 1133 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
FT METAL 1139 1139 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
FT METAL 1144 1144 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
FT NP_BIND 1080 1132 FMN (BY SIMILARITY).
FT CONFLICT 105 109 GWRLA -> LAFR (IN REF. 1).
FT CONFLICT 140 140 E -> N (IN REF. 1).
FT CONFLICT 220 240 LCMTDLPREYLDLADLRLES -> CVCRRICRVLSGSCP
ASGM (IN REF. 1).
FT CONFLICT 259 282 LAQPFYLAHNGEINTITGNROWA -> WNRSAIWRITVK
STSPVTANG (IN REF. 1 AND 3).
FT CONFLICT 460 460 N -> K (IN REF. 1).
FT CONFLICT 574 574 A -> T (IN REF. 1).
FT CONFLICT 863 863 D -> H (IN REF. 1).
FT CONFLICT 1286 1290 THGQQ -> DARRS (IN REF. 1).
FT CONFLICT 1373 1373 G -> A (IN REF. 1).
FT CONFLICT 1376 1376 G -> R (IN REF. 1).
SQ SEQUENCE 1517 AA; 166709 MW; A37F7221 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 1517;
Best Local Similarity 50.0%; Pred. No. 8.44e+00;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 340 WQNPDMDE 349
QY 213 REDNPKNPE 222
:::|||||

RESULT 10
ID NIAM_BOVIN STANDARD; PRT; 186 AA.
AC Q0372;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 13-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE ASHI SUBUNIT PRECURSOR (EC 1.6.5.3)
DE (EC 1.6.99.3) (COMPLEX I-ASHI) (CI-ASHI).
GN NDUFB8.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDEA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=HEART;
RX MEDLINE: 92389317
RA WALKER J.E., ARIZMENDI J.M., DUPUIS A., FEARNEY I.M., FINEL M.,
RA MEDD S.M., PILKINGTON S.J., RUNSWICK M.J., SKEHEL J.M.;
RT "Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from
RT bovine heart mitochondria. Application of a novel strategy for
RT sequencing proteins using the polymerase chain reaction.";
RL J. MOL. BIOL. 226:1051-1072(1992).
CC -|- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC TO BE UBIQUINONE.
CC -|- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -|- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
CC -|- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE; MATRIX SIDE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: X63209; G234; -
DR PIR: S28242; S28242.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION; TRANSIT PEPTIDE.
FT TRANSIT 1 28 MITOCHONDRION.
FT CHAIN 29 186 NADH-UBIQUINONE REDUCTASE ASHI SUBUNIT.
SQ SEQUENCE 186 AA; 21652 MW; ED444866 CRC32;

Query Match 66.7%; Score 44; DB 1; Length 186;
Best Local Similarity 50.0%; Pred. No. 1.38e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 170 RGGDPNKEPE 179
QY 213 REDNPKNPE 222
:::|||||

RESULT 11
ID HMG1_CHICK STANDARD; PRT; 201 AA.
AC P40618;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HIGH MOBILITY GROUP PROTEIN HMG2A (HMG-2A).
GN HMG2A.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN; TISSUE=LIVER;
RX MEDLINE: 92223100.
RA OKA T., ENDO Y., ITO M., MIYAMOTO K.I., SASAKAWA T., SUZUKI I.,
RA NATORI Y.
RT "Molecular cloning of chick liver HMG 2a cDNA and developmental
RT expression of HMG 2a mRNA."
RL BIOCHIM. BIOPHYS. ACTA 1130:224-226(1992).
CC -|- FUNCTION: BINDS PREFERENTIALLY SINGLE-STRANDED DNA AND UNWINDS
CC DOUBLE STRANDED DNA.
CC -|- SUBCELLULAR LOCATION: NUCLEAR.
CC -|- DEVELOPMENTAL STAGE: FOUND IN NEWLY HATCHED CHICK LIVER AND
CC DECREASES DURING POSTNATAL DEVELOPMENT.
CC -|- SIMILARITY: BELONGS TO THE HMG1/HMG2 PROTEIN FAMILY.
CC -|- SIMILARITY: CONTAINS 2 HMG BOXES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: X63463; G63494; -
DR PIR: S22359; S22359.
DR PROSITE: PS00353; HMG1_2; 1.
DR PFAM: PF00505; HMG_box; 2.
DR HSP: P07155; 1AAB.
KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; REPEAT.
FT INIT_MET 0 0 BY SIMILARITY.
FT DNA_BIND 8 78 HMG BOX 1.
FT DNA_BIND 92 160 HMG BOX 2.
SQ SEQUENCE 201 AA; 22940 MW; 910084FD CRC32;

Query Match 66.7%; Score 44; DB 1; Length 201;
Best Local Similarity 60.0%; Pred. No. 1.38e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 23 REHKKKNE 32
QY 213 REDNPKNE 222

Best Local Similarity 60.0%; Pred. No. 1.38e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 118 REDKPEPPE 127
QY 213 REDNPKNE 222

Best Local Similarity 60.0%; Pred. No. 1.38e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 12
ID APX1_PEA STANDARD; PRT; 249 AA.
AC P48534;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE L-ASCORBATE PEROXIDASE, CYTOSOLIC (EC 1.11.1.11) (AP).
GN APX1 OR APX1.
OS PISUM SATIVUM (GARDEN PEA).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC FABALES; FABACEAE; PAPILIONOIDEAE; PISUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, LITTLE MARVEL; TISSUE=LEAF;
RX MEDLINE: 92008636.
RA MITTLER R., ZILINSKAS B.A.;
RT "Molecular cloning and nucleotide sequence analysis of a cDNA
encoding pea cytosolic ascorbate peroxidase.";
RL FEBS LETT. 289:257-259(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, LITTLE MARVEL;
RX MEDLINE: 93016138.
RA MITTLER R., ZILINSKAS B.A.;
RT "Molecular cloning and characterization of a gene encoding pea
cytosolic ascorbate peroxidase.";
RL J. BIOL. CHEM. 267:21802-21807(1992).
RN [3]
RP ERRATUM.
RA MITTLER R., ZILINSKAS B.A.;
RL J. BIOL. CHEM. 268:4568-4568(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RA PATTERSON W.R., POULOS T.L.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: PLAYS A KEY ROLE IN HYDROGEN PEROXIDE REMOVAL IN
CC THE CHLOROPLASTS AND CYTOSOL OF HIGHER PLANTS.
CC -!- CATALYTIC ACTIVITY: L-ASCORBATE + H(2)O(2) = DEHYDROASCORBATE +
CC 2 H(2)O.
CC -!- COFACTOR: HEME.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- INDUCTION: BY STRESS.
CC -!- SIMILARITY: HIGH, TO YEAST CYTOCHROME C PEROXIDASE (P00431).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M93051; G169043; -
CC EMBL: X62077; G20648; -
CC FDB: IAPX; 08-NAR-96.
CC DR PROSITE: PS00435; PEROXIDASE_1; 1.
CC DR PROSITE: PS00436; PEROXIDASE_2; 1.
CC DR PFAM: PF00141; peroxidase; 2.
KW OXIDOREDUCTASE; PEROXIDASE; HYDROGEN PEROXIDE; HEME; 3D-STRUCTURE.
FT INIT_MET 0
FT ACT_SITE 37 37 BY SIMILARITY.
FT ACT_SITE 41 41 DISTAL HISTIDINE (BY SIMILARITY).
FT ACT_SITE 162 162 PROXIMAL HISTIDINE (HEME AXIAL LIGAND)
FT (BY SIMILARITY).
SQ SEQUENCE 249 AA; 27061 MW; 4183ED17 CRC32;
Query Match 66.7%; Score 44; DB 1; Length 249;

Best Local Similarity 60.0%; Pred. No. 1.38e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 118 REDKPEPPE 127
QY 213 REDNPKNE 222

Best Local Similarity 60.0%; Pred. No. 1.38e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 13
ID ATPC_SPIPL STANDARD; PRT; 310 AA.
AC P50006;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ATP SYNTHASE GAMMA CHAIN (EC 3.6.1.34).
GN ATPG OR ATPC.
OS SPIRULINA PLATENSIS.
OC BACTERIA; CYANOBACTERIA; OSCILLATORIALES; SPIRULINA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.
RC STRAIN=CL;
RX MEDLINE: 95337127.
RA STEINEMANN D., LILL H.;
RT "Sequence of the gamma-subunit of Spirulina platensis: a new
principle of thiol modulation of F0F1 ATP synthase";
RL BIOCHIM. BIOPHYS. ACTA 1230:86-90(1995).
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE GAMMA CHAIN IS BELIEVED TO BE
CC IMPORTANT IN REGULATING ATPASE ACTIVITY AND THE FLOW OF PROTONS
CC THROUGH THE CF(0) COMPLEX.
CC -!- ENZYME REGULATION: THIOL-MODULATION BY RAISING THE ENZYME'S
CC ACTIVATION THRESHOLD UPON OXIDATION OF THE CYSTEINES, THEREBY
CC PREVENTING WASTEFUL ATP-HYDROLYSIS.
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -!- SIMILARITY: BELONGS TO THE ATPASE GAMMA CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 246799; G577821; -
CC PROSITE: PS00153; ATPASE_GAMMA; FALSE_NEG.
CC PFAM: PF00231; ATP-synt; 1.
CC DR HSP; P05631; ICOM.
KW ATP SYNTHESIS; CF(1); HYDROGEN ION TRANSPORT; HYDROLASE.
FT INIT_MET 0
FT DISULFID 66 137 POTENTIAL.
SQ SEQUENCE 310 AA; 34361 MW; D1F1857B CRC32;
Query Match 66.7%; Score 44; DB 1; Length 310;
Best Local Similarity 62.5%; Pred. No. 1.38e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 139 DNPDKNE 146
QY 215 DNPDKNE 222

Best Local Similarity 62.5%; Pred. No. 1.38e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

RESULT 14
ID PROB_THETH STANDARD; PRT; 376 AA.
AC Q60050;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GLUTAMATE 5-KINASE (EC 2.7.2.11) (GAMMA-GLUTAMYL KINASE) (GK).
GN PROB.

```
OS THERMUS AQUATICUS (SUBSP. THERMOPHILUS).
OC BACTERIA; THERMUS/DEINOCOCCUS GROUP; THERMUS GROUP; THERMUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN:HB27;
RC MEDLINE: 95080620.
RA KOSUGE T., TABATA K., HOSHINO T.;
RT "Molecular cloning and sequence analysis of the proBA operon from an
RT extremely thermophilic eubacterium thermus thermophilus.";
RL FEMS MICROBIOL. LETT. 123:55-61(1994).
CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE = ADP + L-GLUTAMATE
CC 5-PHOSPHATE (PRODUCT RAPIDLY CYCLIZES TO 5-OXOPROLINE AND
CC ORTHOPHOSPHATE).
CC -!- PATHWAY: FIRST STEP IN PROLINE BIOSYNTHESIS PATHWAY.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GLUTAMATE 5-KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D29973; G577651; -
DR PROSITE; PS00902; GLUTAMATE_5_KINASE; 1.
DR PFAM; PF00696; askinase; 1.
KW TRANSFERASE; KINASE; PROLINE BIOSYNTHESIS.
SQ SEQUENCE 376 AA; 40869 MW; 8151C222 CRC32;

Query Match 66.7%; Score 44; DB 1; Length 376;
Best Local Similarity 55.6%; Pred. No. 1.38e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 184 EEDPKNFKQ 192
QY 214 EDNPKNPE 222
I::I I::I

RESULT 15
ID XPR6_YARLI STANDARD; PRT; 976 AA.
AC P42781;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE DIABASIC PROCESSING ENDOPROTEASE PRECURSOR (EC 3.4.21.-).
GN XPR6.
OS YARROWIA LIPOLYTICA (CANDIDA LIPOLYTICA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC DIPODASCACEAE; YARROWIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:CX161-1B;
RX MEDLINE: 94262316.
RA ENDERLIN C.S., OGRYDZIAK D.M.;
RT "Cloning, nucleotide sequence and functions of XPR6, which codes for
RT a dibasic processing endoprotease from the yeast Yarrowia
RT lipolytica.";
RL YEAST 10:67-79(1994).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY. HIGH SIMILARITY WITH OTHER FURIN-LIKE ENZYMES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L16238; G295683; -
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
```

```
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PFAM; PF00082; subtilase; 1.
DR HSSP; P04072; ITEC.
KW HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; CALCIUM-BINDING;
KW TRANSMEMBRANE; ZYMOGEN; SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 976 DIABASIC PROCESSING ENDOPROTEASE.
FT DOMAIN 211 230 ASP/GLU-RICH (HIGHLY ACIDIC).
FT TRANSMEM 524 544 POTENTIAL.
FT TRANSMEM 855 875 POTENTIAL.
FT ACT_SITE 311 311 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 528 528 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 156 156 POTENTIAL.
FT CARBOHYD 291 291 POTENTIAL.
FT CARBOHYD 299 299 POTENTIAL.
FT CARBOHYD 336 336 POTENTIAL.
FT CARBOHYD 606 606 POTENTIAL.
FT CARBOHYD 886 886 POTENTIAL.
SQ SEQUENCE 976 AA; 110024 MW; 69F2896A CRC32;

Query Match 66.7%; Score 44; DB 1; Length 976;
Best Local Similarity 50.0%; Pred. No. 1.38e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 782 KEDKPEDKPE 791
QY 213 REDNPKNPE 222
::: I::: I::
```

Search completed: Thu Jul 8 18:31:20 1999
Job time : 9 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:31:37 1999; Maspar time 7.84 Seconds
69.579 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (213-222) from US09041236.pep (13 of 45)

Perfect Score: 66
Sequence: 1 REDNPDKNPE 10

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 20.952; Variance 25.252; scale 0.830

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description	Pred. No.
1	66	100.0	393 11	SEMAPHORIN L (FRAGMENT)	2.06e-04
2	66	100.0	075326	SEMAPHORIN L.	2.06e-04
3	50	75.8	037601	NKX5-1 PROTEIN.	1.45e+00
4	49	74.2	273 14	COAT PROTEIN (AA 1 - 2	2.42e+00
5	49	74.2	431 5	R102.5 PROTEIN.	2.42e+00
6	49	74.2	1371 14	3'PART OF GENOME (FRAG	2.42e+00
7	49	74.2	1377 14	POLYPEPTIDE (FRAGMENT)	2.42e+00
8	49	74.2	3056 14	POLYPEPTIDE.	2.42e+00
9	48	72.7	077690	OLFACTORY ENTEROTOXIN	4.01e+00
10	48	72.7	1075 13	GUANYLYL CYCLASE C.	4.01e+00
11	48	72.7	1075 13	MEMBRANE GUANYLYL CYCL	4.01e+00
12	46	69.7	263 4	APK1 ANTIGEN.	1.08e+01
13	45	68.2	67 5	BALBIANI RING (BR6), S	1.75e+01
14	45	68.2	374 5	C24G6.5 PROTEIN.	1.75e+01
15	45	68.2	426 10	DNAJ-LIKE PROTEIN.	1.75e+01
16	45	68.2	439 5	TOSC3.5 PROTEIN.	1.75e+01
17	45	68.2	534 2	IMMUNOGENIC SECRETED P	1.75e+01
18	45	68.2	596 5	KO1G12.3 PROTEIN.	1.75e+01
19	45	68.2	1065 13	GUANYLYL CYCLASE C (EC	1.75e+01
20	45	68.2	1085 2	COBN PROTEIN.	1.75e+01

21	45	68.2	1741 5	046095	COSMID 87B1.	1.75e+01
22	45	68.2	1768 5	Q24153	PARALLEL SISTER CHROMA	1.75e+01
23	45	68.2	2161 4	Q01668	NEURONAL L-TYPE CALCIU	1.75e+01
24	45	68.2	2170 4	Q13931	VOLTAGE-DEPENDENT CALC	1.75e+01
25	45	68.2	2181 4	Q13916	NEUROENDOCRINE/BETA-CE	1.75e+01
26	44	66.7	103 1	Q48241	NON-HISTONE PROTEIN.	2.82e+01
27	44	66.7	200 4	Q15347	HIGH MOBILITY GROUP PR	2.82e+01
28	44	66.7	200 11	Q54879	HIGH MOBILITY GROUP PR	2.82e+01
29	44	66.7	208 13	Q91070	HMG PROTEIN.	2.82e+01
30	44	66.7	249 10	Q96399	CYTOSOLIC ASCORBATE PE	2.82e+01
31	44	66.7	250 10	Q41712	ASCORBATE PEROXIDASE.	2.82e+01
32	44	66.7	250 10	Q39843	ASCORBATE PEROXIDASE 2	2.82e+01
33	44	66.7	250 10	Q48919	ASCORBATE ASCORBATE PE	2.82e+01
34	44	66.7	250 10	Q49159	CYTOSOLIC ASCORBATE PE	2.82e+01
35	44	66.7	250 10	Q43758	ASCORBATE PEROXIDASE (2.82e+01
36	44	66.7	359 6	Q28450	INTERPHOTORECEPTOR RET	2.82e+01
37	44	66.7	446 5	Q21730	R04F11.3 PROTEIN.	2.82e+01
38	44	66.7	466 4	Q15778	P167 (FRAGMENT).	2.82e+01
39	44	66.7	481 5	Q22539	COSMID T17H7.	2.82e+01
40	44	66.7	856 5	Q21186	COSMID K03C7.	2.82e+01
41	44	66.7	1087 5	Q20948	F57C7.1B PROTEIN.	2.82e+01
42	44	66.7	1250 5	Q20947	F57C7.1A PROTEIN.	2.82e+01
43	44	66.7	1382 4	Q14152	K1AA0139 PROTEIN.	2.82e+01
44	44	66.7	1684 2	Q03658	EXO-MALTOPEPTAACHYDROLA	2.82e+01
45	43	65.2	4005 4	Q13744	ALL-1 PROTEIN.	4.50e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	393 AA.
ID	088371			
AC	088371			
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	SEMAPHORIN L (FRAGMENT).			
GN	SEMAL.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 98389619.			
RA	LANGHE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;			
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA			
RT	viruses".			
RL	GENOMICS 51:340-350(1998).			
DR	EMBL; AF030699; G3523117; -.			
FT	NON_TER 393 393			
SQ	SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;			

Query Match 100.0%; Score 66; DB 11; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.06e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	243	REDNPDKNPE 252
Qy	213	REDNPDKNPE 222

RESULT	2	PRELIMINARY;	PRT;	666 AA.
ID	075326			
AC	075326			
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	SEMAPHORIN L.			
GN	SEMAL.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC	CATARRHINI; HOMINIDAE; HOMO.			
RN	[1]			
RP	SEQUENCE FROM N.A.			

RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
viruses";
RL GENOMICS 51:340-350(1998).
DR EMBL: AF030698; G3523115; -.
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 66; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 2.06e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 245 REDNPKNPE 254
QY 213 REDNPKNPE 222

RESULT 3
ID O57601 PRELIMINARY; PRT; 308 AA.
AC O57601;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE NXK5-1 PROTEIN.
GN NXK5-1.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA RINKWITZ-BRANDT S., BOBER E.;
RA HENKWIETZ-HANDT S., BOBER E.;
RA RINKWITZ-H., GUTHRIE S., HADRY S., HOFFMANN S., ARNOLD H.H.,
RL DEVELOPMENT 0-0-0(0).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: Y15989; E1217999; -.
DR PROSITE: PS00027; HOMFOBOX_1: 1.
KW HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN.
SQ SEQUENCE 308 AA; 33787 MW; 67AADB33 CRC32;

Query Match 75.8%; Score 50; DB 13; Length 308;
Best Local Similarity 55.6%; Pred. No. 1.45e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 170 REBSEKPP 178
QY 213 REDNPKNP 221

RESULT 4
ID Q84851 PRELIMINARY; PRT; 273 AA.
AC Q84851;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COAT PROTEIN (AA 1 - 273) (FRAGMENT).
OS BEAN YELLOW MOSAIC VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DANISH GLADIOLUS.;
RX MEDLINE; 90370489.
RA BOYE K., JENSEN P.E., STUMMANN B.M., HENNINGSEN K.W.;
RT "Nucleotide sequence of cDNA encoding the BYMV coat protein gene";
RL NUCLEIC-ACIDS RES. 18:4926-4926(1990).
DR EMBL: X53684; G61218; -.
DR PFAM: PF00767; Poty_coat; 1.
KW COAT PROTEIN.
FT NON_TER 1 1
SQ SEQUENCE 273 AA; 30789 MW; D1159F30 CRC32;

Query Match 74.2%; Score 49; DB 14; Length 273;

Best Local Similarity 77.8%; Pred. No. 2.42e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 20 EENPDKNSE 28
QY 214 EENPDKNPE 222

RESULT 5
ID Q21891 PRELIMINARY; PRT; 431 AA.
AC Q21891;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE R102.5 PROTEIN.
GN R102.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA BERKS M.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.

RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL NATURE 368:32-38(1994).
DR EMBL: 270309; E1348680; -.
SQ SEQUENCE 431 AA; 48907 MW; 6ED4A59C CRC32;

Query Match 74.2%; Score 49; DB 5; Length 431;
Best Local Similarity 40.0%; Pred. No. 2.42e+00;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 110 KOEDPDRTPE 119
QY 213 REDNPKNPE 222

RESULT 6
ID Q65883 PRELIMINARY; PRT; 1371 AA.
AC Q65883;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 3'PART OF GENOME (FRAGMENT).
OS BEAN YELLOW MOSAIC VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GLADIOLUS;
RX MEDLINE; 92288317.
RA BOYE K., STUMMANN B.M., HENNINGSEN K.W.;
RT "cDNA cloning and sequencing of the bean yellow mosaic virus nuclear
inclusion protein genes";
RL PLANT MOL. BIOL. 18:1203-1205(1992).
DR EMBL: X63358; G312734; -.
DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
DR PFAM: PF00767; Poty_coat; 1.
DR PFAM: PF00863; Peptidase_C4; 1.

```
KW COAT PROTEIN; POLYPROTEIN; PROTEASE.
FT NON_TER 1 1
FT CHAIN 1 92 POTENTIAL.
FT CHAIN 93 145 POTENTIAL.
FT CHAIN 146 579 POTENTIAL.
FT CHAIN 580 1098 POTENTIAL.
FT CHAIN 1099 1371 POTENTIAL.
SQ SEQUENCE 1371 AA; 155447 MW; 7D894BE7 CRC32;

Query Match 74.2%; Score 49; DB 14; Length 1371;
Best Local Similarity 77.8%; Pred. No. 2.42e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1118 EENPDKNSE 1126
QY 214 EDNPDKNPE 222

RESULT 7
ID Q83137 PRELIMINARY; PRT; 1377 AA.
AC Q83137;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE POLYPEPTIDE (FRAGMENT).
OS BEAN YELLOW MOSAIC VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MB4;
RA NAKAMURA S., HONKURA R., IWAI T., UGAKI M., OHASHI M., OHASHI Y.;
RT "Nucleotide sequence of the 3'-terminal region of bean yellow mosaic
virus RNA and resistance to viral infection in transgenic Nicotiana
benthiana expressing its coat protein gene.";
RL ANN. PHYTOPATHOL. SOC. JPN. 60:295-304(1994).
DR EMBL; D28819; D1006530; -
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poty_coat; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
KW COAT PROTEIN.
FT NON_TER 1 98 CI.
FT CHAIN 99 151 6K.
FT CHAIN 152 585 NIA.
FT CHAIN 586 1104 NIB.
FT CHAIN 1105 1377 COAT PROTEIN.
SQ SEQUENCE 1377 AA; 156192 MW; 352C1E92 CRC32;

Query Match 74.2%; Score 49; DB 14; Length 1377;
Best Local Similarity 77.8%; Pred. No. 2.42e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1124 EENPDKNSE 1132
QY 214 EDNPDKNPE 222

RESULT 8
ID Q65892 PRELIMINARY; PRT; 3056 AA.
AC Q65892;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS BEAN YELLOW MOSAIC VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MB4;
RA NAKAMURA S.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
```

```
RN SEQUENCE OF 1679-3056 FROM N.A.
RP STRAIN-MB4;
RA NAKAMURA S., HONKURA R., UGAKI M., OHASHI M., OHASHI Y.;
RT "Nucleotide sequence of the 3'-terminal region of bean yellow mosaic
virus RNA and resistance to viral infection in transgenic Nicotiana
benthiana expressing its coat protein gene.";
RL ANN. PHYTOPATHOL. SOC. JPN. 60:295-304(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-MB4;
RA NAKAMURA S., HONKURA R., IWAI T., UGAKI M., OHASHI Y.;
RT "The complete nucleotide sequence of bean yellow mosaic virus genomic
RNA.";
RL ANN. PHYTOPATHOL. SOC. JPN. 62:472-477(1996).
DR EMBL; D83749; D1012773; -
DR PFAM; PF00271; helicase_C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poty_coat; 1.
DR PFAM; PF00851; Peptidase_C6; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
KW POLYPROTEIN.
SQ SEQUENCE 3056 AA; 347572 MW; 469E4459 CRC32;

Query Match 74.2%; Score 49; DB 14; Length 3056;
Best Local Similarity 77.8%; Pred. No. 2.42e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 2803 EENPDKNSE 2811
QY 214 EDNPDKNPE 222

RESULT 9
ID O77690 PRELIMINARY; PRT; 1072 AA.
AC O77690;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE OLFACTORY ENTEROTOXIN RECEPTOR.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RA GORACZNIK R.M., DUDA T., SHARMA R.K.;
RT "Cloning of Enterotoxin Receptor Guanylate Cyclase (Sta-RGC) from
Bovine Olfactory.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF081464; G3411272; -
DR PROSITE; PS00452; GUANYLATE_CYCLASES; 1.
KW LYASE.
SQ SEQUENCE 1072 AA; 122709 MW; DBEE56C5 CRC32;

Query Match 72.7%; Score 48; DB 6; Length 1072;
Best Local Similarity 40.0%; Pred. No. 4.01e+00;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 727 WEEDPEKRPD 736
QY 213 REDNPDKNPE 222

RESULT 10
ID O93252 PRELIMINARY; PRT; 1075 AA.
AC O93252;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GUANYLYL CYCLASE C.
GN OLCG6.
OS ORYZIAS LATIPES (MEDAKA FISH).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
```

```

OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; ATHERINOMORPHA;
OC CYPRINODONTIFORMES; ADRIANICHTHYOIDEI; ADRIANICHTHYIDAE; ORYZIINAE;
OC ORYZIAS.
RN
RP SEQUENCE FROM N.A.
RA KUSAKABE T., MANTOKU T., SUZUKI N.;
RT "Gene encoding medaka fish guanylyl cyclase C.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB016081; D1032669; -.
KW PROSITE; PS00452; GUANYLATE_CYCLASES; 1.
KW LYASE.
SQ SEQUENCE 1075 AA; 122847 MW; 67ED6778 CRC32;

Query Match 72.7%; Score 48; DB 13; Length 1075;
Best Local Similarity 40.0%; Pred. No. 4.01e+00;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 718 WEEDPEKRPD 727
QY 213 REDNPKNPE 222
:|:|:|:|:

RESULT 11
ID O42129 PRELIMINARY; PRT; 1075 AA.
AC O42129;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1999 (TREMBLREL. 09, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST SEQUENCE UPDATE)
DE MEMBRANE GUANYLYL CYCLASE.
GN OLCG6
OS ORYZIAS LATIPES (MEDAKA FISH).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; ATHERINOMORPHA;
OC CYPRINODONTIFORMES; ADRIANICHTHYOIDEI; ADRIANICHTHYIDAE; ORYZIINAE;
OC ORYZIAS.
RN
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE.
RA MANTOKU T., SUZUKI N.;
RT "mcdaka ( Oryzias latipes ) membrane guanylyl cyclase OLCG6.";
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB007192; D1034924; -.
KW PROSITE; PS00452; GUANYLATE_CYCLASES; 1.
KW MEMBRANE; LYASE.
SQ SEQUENCE 1075 AA; 122788 MW; 454DCDE3 CRC32;

Query Match 72.7%; Score 48; DB 13; Length 1075;
Best Local Similarity 40.0%; Pred. No. 4.01e+00;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 718 WEEDPEKRPD 727
QY 213 REDNPKNPE 222
:|:|:|:|:

RESULT 12
ID Q16206 PRELIMINARY; PRT; 263 AA.
AC Q16206;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE AP1 ANTIGEN.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN
RP SEQUENCE FROM N.A.
RA CHANG K., PASTAN I.;
RX MEDLINE; 94200897.
RT "Molecular cloning and expression of a cDNA encoding a protein
RT detected by the Ki antibody from an ovarian carcinoma (OVCAR-3) cell
RL line.";
RL INT. J. CANCER 57:90-97(1994).

DR EMBL; S72904; G633926; -.
SQ SEQUENCE 263 AA; 30530 MW; 763858A3 CRC32;

Query Match 69.7%; Score 46; DB 4; Length 263;
Best Local Similarity 50.0%; Pred. No. 1.08e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 97 REDDPNKEQQ 106
QY 213 REDNPKNPE 222
||:|:|:|:

RESULT 13
ID Q23779 PRELIMINARY; PRT; 67 AA.
AC Q23779;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE BALBIANI RING (BR6), SECRETORY PROTEIN SP-IC REPETITIVE REGION
DE (FRAGMENT).
OS CHIRONOMUS PALLIDIVITTATUS (MIDGE).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; NEMATOCERA; CHIRONOMOIDEA; CHIRONOMIDAE;
OC CHIRONOMINAE; CHIRONOMUS.
RN
RP SEQUENCE FROM N.A.
RA GALLER R., RYLANDER L., RIEDEL N., KLUDING H., EDSTROM J.E.;
RL PROC. NATL. ACAD. SCI. U.S.A. 81:1448-1452(1984).
DR EMBL; K01695; G156548; -.
FT NON_TER 1
FT NON_TER 67
SQ SEQUENCE 67 AA; 8408 MW; 0A39B395 CRC32;

Query Match 68.2%; Score 45; DB 5; Length 67;
Best Local Similarity 40.0%; Pred. No. 1.75e+01;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 57 RPEEPEREPE 66
QY 213 REDNPKNPE 222
|:|:|:|:|:

RESULT 14
ID O76381 PRELIMINARY; PRT; 374 AA.
AC O76381;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE C24G6.5 PROTEIN.
GN C24G6.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., LAISTER N., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTON N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RL NATURE 368:32-38(1994).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
```

RA GRECO T., BRADSHAW H., KEPPLER D.;
RT "The sequence of C. elegans cosmid C24G6."
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -|- SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC
CC DNAJ-LIKE PROTEINS.
DR EMBL; AF067936; G3193178; -;
DR PROSITE; PS00637; DNAJ_CXXCXGXG; 1.
KW CHAPERONE; DNA REPLICATION.
SQ SEQUENCE 374 AA; 40683 MW; 5202F456 CRC32;

Query Match 68.2%; Score 45; DB 5; Length 374;
Best Local Similarity 60.0%; Pred. No. 1.75e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 122 KEYHPDKNPD 131
QY 213 REDNPKNPE 222

RESULT 15
ID 004662 PRELIMINARY; PRT; 426 AA.
AC 004662;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE DNAJ-LIKE PROTEIN.
GN J10.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RA LIN W., LIN B.L.;
RL PLANT PHYSIOL. 113:863-863(1997).
DR EMBL; Y11969; E1169949; -;
DR PROSITE; PS00636; DNAJ_1; 1.
DR PFAM; PF00226; DnaJ_1;
DR MENDEL; 16533; ARATH; 2537; mnl16533.
SQ SEQUENCE 426 AA; 48210 MW; 31EF988C CRC32;

Query Match 68.2%; Score 45; DB 10; Length 426;
Best Local Similarity 60.0%; Pred. No. 1.75e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 31 QVHPDKNPN 40
QY 213 REDNPKNPE 222

Search completed: Thu Jul 8 18:31:50 1999
Job time : 13 secs.

THIS PAGE BLANK (USPTO)

WQSELA

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:34:39 1999; MasPar time 11.04 Seconds
Tabular output not generated.
17.335 Million cell updates/sec

Title: >US-09-041-236-2
Description: (222-230) from US09041236.pep (14 of 45)
Perfect Score: 58
Sequence: 1 EAPLNVSrv 9
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 2126608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 15.753; Variance 39.699; scale 0.397

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	48	82.8	229	8 R41729	Peptide derived from	3.49e+01
2	48	82.8	545	8 R41726	Peptide derived from	3.49e+01
3	47	81.0	413	23 W23649	Recombinant squirrel	4.68e+01
4	46	79.3	329	32 W38396	Homo sapiens HLIW-2 p	6.26e+01
5	45	77.6	661	5 R26176	GORGab protein.	8.36e+01
6	45	77.6	661	2 R22674	Recombinant GOR gab p	8.36e+01
7	44	75.9	347	14 R61990	Porphyromonas gengiva	1.11e+02
8	44	75.9	348	14 R61991	Porphyromonas gengiva	1.11e+02
9	44	75.9	348	14 R61989	Porphyromonas gengiva	1.11e+02
10	44	75.9	350	14 R61992	Porphyromonas gengiva	1.11e+02
11	43	74.1	360	14 R61994	Porphyromonas gengiva	1.48e+02
12	43	74.1	597	22 W20536	H. pylori chaperone p	1.48e+02
13	43	74.1	638	22 W21012	H. pylori chaperone p	1.48e+02
14	43	74.1	982	35 W77288	Zebrafish differentia	1.48e+02
15	42	72.4	413	23 W23652	Recombinant squirrel	1.96e+02
16	41	70.7	293	36 W69419	Protein encoded by in	2.59e+02

17	41	70.7	309	1 P81324	Homoserine kinase.	2.59e+02
18	41	70.7	309	3 P61775	Homoserine kinase.	2.59e+02
19	41	70.7	359	34 W61491	Human fetuin glycopro	2.59e+02
20	41	70.7	382	26 W31910	Human MAPKAP kinase 3	2.59e+02
21	41	70.7	513	8 R40187	Solanum melongena Fla	2.59e+02
22	41	70.7	560	25 W35382	Murine metastatic nuc	2.59e+02
23	41	70.7	620	6 R32548	Tomato hsp80.	2.59e+02
24	41	70.7	656	22 W11216	Leishmania braziliens	2.59e+02
25	41	70.7	656	35 W70211	Leishmania antigen Lb	2.59e+02
26	41	70.7	699	6 R32549	Tomato hsp80 encoded	2.59e+02
27	41	70.7	704	1 P80087	Sequence of 85 kd pro	2.59e+02
28	41	70.7	802	16 R90848	Gibberellin (GAL), en	2.59e+02
29	41	70.7	817	7 R34942	Human epidermal trans	2.59e+02
30	40	69.0	152	16 R88603	Adometase (I2V) SAM-K	3.42e+02
31	40	69.0	152	25 W29147	S-adenosylmethionine	3.42e+02
32	40	69.0	152	17 R88606	S-adenosylmethionine	3.42e+02
33	40	69.0	277	10 R54203	snab gene product inv	3.42e+02
34	40	69.0	413	24 W23138	Protease inhibitor	3.42e+02
35	40	69.0	445	29 W47399	B. lactofermentum dia	3.42e+02
36	40	69.0	445	24 W32285	Brevibacterium lactof	3.42e+02
37	40	69.0	445	34 W68150	Diaminopimelate decar	3.42e+02
38	40	69.0	445	33 W69552	Brevibacterium lactof	3.42e+02
39	40	69.0	445	14 R72451	Brevibacterium flavum	3.42e+02
40	40	69.0	632	16 R85300	Arabidopsis pathogen	3.42e+02
41	40	69.0	701	32 W48740	Human placental-deriv	3.42e+02
42	40	69.0	917	38 W73512	Rat ICAM-4 protein se	3.42e+02
43	40	69.0	917	32 W60160	Rat intercellular adh	3.42e+02
44	40	69.0	3639	8 R40227	ACVS.	3.42e+02
45	40	69.0	3712	3 R13896	ACV synthetase.	3.42e+02

ALIGNMENTS

RESULT 1
ID R41729 standard; Protein; 229 AA.
AC R41729; 1994 (first entry)
DE Peptide derived from HMW2 gene cluster.
KW HMW; high molecular weight protein; virus; vaccine; influenza;
epitope; immunity; haemophilus influenzae; gene cluster.
PS Haemophilus influenzae.
OS W09319090-A.
PD 30-SEP-1993.
PF 16-MAR-1993; U02166.
PR 16-MAR-1992; GB-005704.
PA (BARE/) BARENKAMP S J.
PI (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI Barenkamp SJ.
DR WFI; 93-320883/40.
DR N-PSDB; Q49509.
PT High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
PS Disclosure: Figure 7: 100pp; English.
CC The isolation and purification of a high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines. This polypeptide is encoded by a
CC sequence found downstream of high molecular weight protein 2 (HMW2).
CC in the gene cluster and is thought to affect the expression of HMW2.
SQ Sequence 229 AA;
Query Match 82.8%; Score 48; DB 8; Length 229;
Best Local Similarity 66.7%; Pred. No. 3.49e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 158 enplkvtrv 166

QY 222 EAPLNVSrv 230

RESULT 2
ID R41726 standard; Protein; 545 AA.

```

AC R41726;
DE 26-APR-1994 (first entry)
DE Peptide derived from HMW1 gene cluster.
KW HMW; high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae; gene cluster.
OS Haemophilus influenzae.
PN W09319090-A.
PD 30-SEP-1993.
PF 16-MAR-1993; U02166.
PR 16-MAR-1992; GB-005704.
PA (BARE/) BARENKAMP S J.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI Barenkamp SJ;
DR WPI; 93-320883/40.
DR N-PSDB; Q49508.
PT High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
PS Disclosure; Figure 6; 100pp; English.
CC The isolation and purification of a high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines. This polypeptide is encoded by a
CC sequence found downstream of high molecular weight protein 1 (HMW1)
CC in the gene cluster and is thought to affect the expression of HMW1.
SQ Sequence 545 AA;

Query Match 82.8%; Score 48; DB 8; Length 545;
Best Local Similarity 66.7%; Pred. No. 3.49e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 158 enplkvtrv 166
| |||:|
QY 222 EAPLNVSRV 230

RESULT 3
ID W23649 standard; Protein; 413 AA.
AC W23649;
DE 08-OCT-1997 (first entry)
DE Recombinant squirrel protease inhibitor HP-55.
DE Protease inhibitor; alpha-1 antitrypsin; HP-55; striped squirrel.
OS Tasmania asiaticus.
FH Key Location/Qualifiers
FT peptide 1..24
FT /label_ Signal
FT protein 25..413
FT /label_ Protease_inhibitor_HP-55
FT region 373..382
FT /label_ Reaction_site
FT J09157299-A.
PN 17-JUN-1997.
PD 04-DEC-1995; 315689.
PF 04-DEC-1995; JP-315689.
PR (KOND/) KONDO J.
PA (KANA-) ZH KANAGAWA KAGAKU GIJUTSU ACAD.
DR WPI; 97-369474/34.
DR N-PSDB; T78176.
PT Squirrel derived protease inhibitor with similarity to alpha-1
PT antitrypsin - shows specific protease activity
PS Claim 3; Page 8-9; 10pp; Japanese.
CC The present sequence represents a recombinant squirrel protease
CC inhibitor protein HP-55 that contains a specifically claimed reaction
CC site, see features table. The protease inhibitor HP-55 has inhibitory
CC activity against elastase but not trypsin or thrombin. The protease
CC inhibitor has a new structure similar to alpha-1-antitrypsin and shows
CC selective inhibitory activity against proteases.
SQ Sequence 413 AA;

Query Match 81.0%; Score 47; DB 23; Length 413;
Best Local Similarity 66.7%; Pred. No. 4.69e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 343 depbtvskv 351
| ||| |||
QY 222 EAPLNVSRV 230

RESULT 4
ID W58396 standard; Protein; 329 AA.
AC W58396;
DE 11-SEP-1998 (first entry)
DE Homo sapiens HLIM-2 protein.
KW HLIM; HLIM-2; LIM protein; prevention; treatment; disease;
KW proliferation; regeneration; cell; tissue; organ; nerve; pancreas;
KW epithelium; accidental injury; degeneration; skin grafts;
KW cancer; arteriosclerosis; abnormal cell growth.
OS Homo sapiens.
PN W09818832-A2.
PD 07-MAY-1998.
PF 10-OCT-1997; U18717.
PR 28-OCT-1996; US-739485.
PA (INCY-) INCYTE PHARM INC.
PA Bandman O, Goli SK, Hillman JL;
WPI: 98-272137/24.
DR N-PSDB; V30951.
DR Human LIM proteins - useful for treating and diagnosing cancer and
PT other proliferative diseases
PS Claim 18; Page 53-54; 94pp; English.
CC The sequence is that of the human LIM protein HLIM-2.
CC The protein, optionally expressed from gene therapy vectors,
CC can be used to treat or prevent cancer or arteriosclerosis or other
CC conditions involving abnormal regulation of cell growth.
CC Also antagonists of it, antisense molecules, ribozymes or
CC triplex-forming agents can be used to induce proliferation and
CC regeneration of cells, tissues and organs (e.g. nerves, pancreas
CC or epithelium), particularly in cases of accidental injury, or
CC degeneration, especially in context of skin grafts. Probes based
CC on the gene sequence are used to detect HLIM-encoding sequences by
CC hybridisation, optionally after amplification, for diagnosis or
CC monitoring of disease, or for assessing predisposition to disease.
CC They may also be used to map the corresponding genomic sequence.
CC Similar diagnostic assays are done at the protein level using
CC antibodies as immunoassay reagents. Antibodies raised against
CC HLIM can also be used therapeutically; for targeting other active
CC agents to cells that express HLIM; in drug screening and to isolate
CC natural HLIM.
SQ Sequence 329 AA;

Query Match 79.3%; Score 46; DB 32; Length 329;
Best Local Similarity 66.7%; Pred. No. 6.26e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 25 eqplaisrv 33
| ||| |||
QY 222 EAPLNVSRV 230

RESULT 5
ID R26176 standard; Protein; 661 AA.
AC R26176;
DE 05-FEB-1993 (first entry)
DE GORgab protein.
KW NANBH virus; non-A non-B hepatitis; detection.
OS Non-A non-B hepatitis virus.
FH Key Location/Qualifiers
FT region 569..623
FT /note_ "GOR 47-1"
FT region 569..597
FT /note_ "spGOR1"
FT region 583..609
FT /note_ "spGOR2"
FT region 597..622
FT /note_ "spGOR3"
FT region 346..363
FT /note_ "spGORb-ep1"

```



```

Db 124 dtpleikrv 132
   :|||: ||
QY 222 EAPLNVSrv 230

RESULT 9
ID R61989 standard; Protein: 348 AA.
AC R61989;
DE 12-DEC-1995 (first entry)
DE Porphyromonas gingivalis strain HW24D-1 fimbrillin protein.
KW Fimbrillin gene; Porphyromonas gingivalis; antibody; diagnosis; primer;
KW Periodontal disease; PCR; amplification.
OS Porphyromonas gingivalis strain HW24D-1.
FH Key Location/Qualifiers
FT peptide 1..10
FT peptide /label= signal peptide
FT peptide 11..348
FT peptide /label= mature peptide
PN WO9510612-A1.
PD 20-APR-1995.
PF 07-OCT-1994; J01687.
PR 08-OCT-1993; JP-253203.
PA (LIOY ) LION CORP.
PI Morishima S;
PI WPI: 95-161797/21.
DR N-PSDB: Q87984.
DR Recombinant fimbrillin protein of Porphyromonas gingivalis - and
PT antibody recognising it, for treatment and diagnosis of periodontal
PT disease
PS Claim 8; Page 52-54; 83pp; Japanese.
CC The amino acid sequence of the fimbrillin protein encoded by the fimA
CC gene from Porphyromonas gingivalis strain HW24D-1. The fimA gene
CC sequence from strain HW24D-1 was obtained by hybridisation with a
CC probe obtained by PCR amplification of the fimA gene from P.gingivalis
CC strain 381 (Q87981) using the primers Q87990-1. The probe was also used
CC to obtain the fimA genes from other P.gingivalis strains (see Q87982-89).
CC The DNAs and antibodies to the proteins, or derived peptides
CC (R61995-R62006) can be used in the detection of P.gingivalis for the
CC diagnosis of periodontal disease.
SQ Sequence 348 AA;

Query Match 75.9%; Score 44; DB 14; Length 348;
Best Local Similarity 44.4%; Pred. No. 1.11e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 124 dtpleikrv 132
   :|||: ||
QY 222 EAPLNVSrv 230

RESULT 10
ID R61992 standard; Protein: 350 AA.
AC R61992;
DE 12-DEC-1995 (first entry)
DE Porphyromonas gingivalis strain ATCC-49417 fimbrillin protein.
KW Fimbrillin gene; Porphyromonas gingivalis; antibody; diagnosis; primer;
KW Periodontal disease; PCR; amplification.
OS Porphyromonas gingivalis strain ATCC-49417.
FH Key Location/Qualifiers
FT peptide 1..10
FT peptide /label= signal peptide
FT peptide 11..350
FT peptide /label= mature peptide
PN WO9510612-A1.
PD 20-APR-1995.
PF 07-OCT-1994; J01687.
PR 08-OCT-1993; JP-253203.
PA (LIOY ) LION CORP.
PI Morishima S;
PI WPI: 95-161797/21.
DR N-PSDB: Q87987.
DR Recombinant fimbrillin protein of Porphyromonas gingivalis - and
PT antibody recognising it, for treatment and diagnosis of periodontal
PT disease
PS Claim 8; Page 60-62; 83pp; Japanese.
CC The amino acid sequence of the fimbrillin protein encoded by the fimA
CC gene from Porphyromonas gingivalis strain HG564. The fimA gene sequence
CC from strain HG564 was obtained by hybridisation with a probe obtained by
CC PCR amplification of the fimA gene from P.gingivalis strain 381 (Q87981)
CC using the primers Q87990-1. The probe was also used to obtain the fimA
CC genes from other P.gingivalis strains (see Q87982-89). The DNAs and
CC antibodies to the proteins, or derived peptides (R61995-R62006) can be
CC used in the detection of P.gingivalis for the diagnosis of periodontal
CC disease.
SQ Sequence 350 AA;

Query Match 74.1%; Score 43; DB 14; Length 360;
Best Local Similarity 62.5%; Pred. No. 1.48e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 138 tplavtrv 145
   :|||: ||
QY 223 APLNVSrv 230

RESULT 12
ID W20536 standard; Protein: 597 AA.
AC W20536;
DT 15-JUL-1997 (first entry)
DE H. pylori chaperone protein 50253.aa.

```

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR N-PSDB: T67868.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptides) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 61; Page 694-695; 1481pp; English.
 CC This sequence is a H. pylori chaperone protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 597 AA;

Query Match 74.1%; Score 43; DB 22; Length 597;
 Best Local Similarity 75.0%; Pred. NO. 1.48e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 334 dlpnvsvr 341

QY 222 EAPLNVSR 229

RESULT 13
 ID W21012 standard; protein; 638 AA.

AC W21012;
 DE 21-JUL-1997 (first entry)
 DE H. pylori chaperone protein, hp5el1211orf10.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR N-PSDB: T68265.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptides) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 61; Page 1401-1402; 1481pp; English.
 CC The present sequence is a Helicobacter pylori chaperone protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds, useful
 CC as potential H. pylori life cycle activators or inhibitors. The genomic
 CC sequence of H. pylori (ATCC 55679) was determined from overlapping
 CC contigs generated by mechanically shearing the bacterial DNA. The
 CC sequences were analysed for ORF of at least 180 nucleotides, and the
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid

CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 638 AA;

Query Match 74.1%; Score 43; DB 22; Length 638;
 Best Local Similarity 75.0%; Pred. NO. 1.48e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 351 dlpnvsvr 358

QY 222 EAPLNVSR 229

RESULT 14
 ID W77288 standard; Protein; 982 AA.

AC W77288;
 DE 20-NOV-1998 (first entry)
 DE Zebrafish differentiation enhancing factor 2 protein.
 KW Zebrafish; differentiation enhancing factor; ankyrin repeat; C2 domain;
 KW SH3 consensus binding sequence; pleckstrin homology domain; adipogenesis;
 KW neurogenesis; hyperplastic disease; neoplastic disease; nervous system.
 OS Brachydanio rerio.
 PN W09836065-A1.
 PD 20-AUG-1998.
 PF 13-FEB-1998; U02724.
 PR 14-FEB-1997; US-038191.
 PA (DAND) DANA FARMER CANCER INST INC.
 PI Chan J, Harris DF, Hu E, King FJ, Spiegelman B,
 PI Thomas RM;
 DR WPI: 98-467173/40.
 DR N-PSDB: V59105.
 PT New nucleic acid encoding differentiation enhancing factor - used
 PT particularly to regulate adipogenesis and neurogenesis, e.g. for
 PT treating tumours and neurological disease
 PS Claim 3; Fig 12; 203pp; English.
 CC The differentiation enhancing factors (DEF), comprise at least one each
 CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology
 CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they
 CC are mediators of SH3-domain dependent signalling and may be involved in
 CC cellular gene expression, cytoskeletal architecture, protein trafficking,
 CC endocytosis or adhesion, migration, proliferation and differentiation of
 CC cells. Typical applications of DEF and its ligand, or of nucleic acid
 CC interaction between the protein and its ligand, or of nucleic acid
 CC expressing them, are treatment of hyperplastic and neoplastic disease
 CC (a wide range of solid tumours and leukaemias), including metastases; for
 CC in vitro induction of differentiation of neural crest cells to neurons,
 CC glial cells etc.; for increasing neuron survival, and inducing cell
 CC repair, in the nervous system (e.g. treatment of traumatic injury,
 CC stroke, Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic
 CC lateral sclerosis, multiple sclerosis etc.).
 SQ Sequence 982 AA;

Query Match 74.1%; Score 43; DB 35; Length 982;
 Best Local Similarity 50.0%; Pred. NO. 1.48e+02;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 655 etplditar 662

QY 222 EAPLNVSR 229

RESULT 15
 ID W23652 standard; Protein; 413 AA.

AC W23652;
 DE 09-OCT-1997 (first entry)
 DE Recombinant squirrel protease inhibitor 55RS.
 KW Protease inhibitor; alpha-1 antitrypsin; Hp-55; striped squirrel;
 KW hibernation related protein.
 OS Tamias asiaticus.
 FH Key Location/Qualifiers

```
FT peptide 1..24
FT /label=" Signal
FT protein 25..413
FT /label=" Protease_inhibitor_55RS
FT region 372..381
FT /label=" Reaction_site
PN J09157298-A.
PD 17-JUN-1997.
PF 04-DEC-1995; 315692.
PR 04-DEC-1995; JP-315692.
PA (KOND/) KONDO J.
PA (KANA-) ZH KANAGAWA KAGAKU GIJUTSU ACAD.
DR WPI; 97-369473/34.
DR N-PSDB; T78180.
PT Squirrel derived protease inhibitor has similar structure to alpha-1
PT antitrypsin - shows selective inhibitory activity against
PT protease(s)
PS Claim 4; Page 8-9; 10pp; Japanese.
CC The present sequence represents a recombinant squirrel protease
CC inhibitor protein 55RS that contains a specifically claimed reaction
CC site, see features table. The protease inhibitor 55RS has inhibitory
CC activity against trypsin, elastase or thrombin. The protease inhibitor
CC has a new structure similar to alpha-1-antitrypsin and shows selective
CC inhibitory activity against proteases. The protease inhibitor cDNA was
CC homologous to cDNA encoding a similar squirrel protein designated
CC hibernation-related protein (HP-55).
SQ Sequence 413 AA;

Query Match 72.4%; Score 42; DB 23; Length 413;
Best Local Similarity 75.0%; Pred. No. 1.96e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 343 eaplsvsk 350
QY 222 EAPLNSR 229
|||||:

Search completed: Thu Jul 8 18:34:58 1999
Job time : 19 secs.
```

M E R E F H

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:34:11 1999; MasPar time 4.98 Seconds
Tabular output not generated. 72.395 Million cell updates/sec

Title: >US-09-041-236-2
Description: (222-230) from US09041236.pep (14 of 45)
Perfect Score: 58
Sequence: 1 EAPLNVS RV 9

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.013; Variance 23.788; scale 0.925

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match %	Length	ID	Description	Pred. No.	
1	47	81.0	521	1	VGIVDH	4.13e+00	
2	46	79.3	327	2	JC4385	6.98e+00	
3	46	79.3	383	2	A36117	6.98e+00	
4	45	77.6	253	2	E69497	1.17e+01	
5	45	77.6	330	2	S71828	1.17e+01	
6	45	77.6	658	2	S76909	1.17e+01	
7	44	75.9	103	1	R5B524	1.94e+01	
8	44	75.9	243	2	S11226	1.94e+01	
9	44	75.9	348	2	JN0918	1.94e+01	
10	44	75.9	348	2	JN0916	1.94e+01	
11	44	75.9	350	2	JN0919	1.94e+01	
12	44	75.9	681	2	E70942	1.94e+01	
13	44	75.9	690	2	H71237	1.94e+01	
14	43	74.1	110	2	S43192	3.20e+01	
15	43	74.1	111	2	E42645	3.20e+01	
16	43	74.1	177	2	A39414	3.20e+01	
17	43	74.1	305	2	D69199	3.20e+01	
18	43	74.1	329	2	S49490	3.20e+01	
19	43	74.1	360	2	JN0921	3.20e+01	
20	43	74.1	360	2	S10512	3.20e+01	
21	43	74.1	372	2	JC2204	3.20e+01	
22	43	74.1	385	2	S78100	3.20e+01	
23	43	74.1	396	2	S39793	3.20e+01	

24 43 74.1 400 2 A26258 endoplasmin - golden 3.20e+01
25 43 74.1 413 2 I56481 alpha 1-proteinase in 3.20e+01
26 43 74.1 442 2 E69997 nitrilotriacetate mon 3.20e+01
27 43 74.1 621 2 B64546 chaperone and heat sh 3.20e+01
28 43 74.1 621 2 A71961 90Kda chaperone - Hel 3.20e+01
29 43 74.1 624 1 HHEC62 heat shock protein C6 3.20e+01
30 43 74.1 626 2 JC5230 heat shock protein ht 3.20e+01
31 43 74.1 631 2 G64048 heat shock protein ht 3.20e+01
32 43 74.1 639 2 F71258 probable heat shock p 3.20e+01
33 43 74.1 650 2 G70169 heat shock protein 90 3.20e+01
34 43 74.1 781 2 S49340 heat-shock protein, 8 3.20e+01
35 43 74.1 795 1 HHCH08 heat shock protein 10 3.20e+01
36 43 74.1 795 2 I50255 108K heat shock prote 3.20e+01
37 43 74.1 802 2 A29317 endoplasmic reticulum 3.20e+01
38 43 74.1 803 2 A35954 endoplasmin precursor 3.20e+01
39 43 74.1 804 2 A53211 glucose-regulated pro 3.20e+01
40 43 74.1 804 2 S51358 protein kinase ppk98 3.20e+01
41 43 74.1 1142 2 S36845 myosin-binding protei 3.20e+01
42 43 74.1 1429 2 JN0609 nitric-oxide synthase 3.20e+01
43 43 74.1 1530 2 S01393 DNA-directed RNA poly 3.20e+01
44 43 74.1 2470 2 I50726 cation-independent ma 3.20e+01
45 43 74.1 4725 1 A44357 dynein heavy chain, c 3.20e+01

ALIGNMENTS

RESULT 1
ENTRY VGIVDH #type complete
TITLE envelope glycoprotein - Dhori virus (strain India/1313/61)
ORGANISM #formal_name Dhori virus
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 05-Sep-1997

ACCESSIONS A34679
REFERENCE A34679
#authors Freedman-Faulstich, E.Z.; Fuller, F.J.
#journal Virology (1990) 175:10-18
#title Nucleotide sequence of the tick-borne, orthomyxo-like Dhori/Indian/1313/61 virus envelope gene.
#cross-references MUID:90177204
#accession A34679
#molecule_type mRNA
#residues 1-571 #label FRE
#cross-references EMBL:M34002; NID:g335189; PID:g335190

GENETICS
#map_position segment 4
CLASSIFICATION #superfamily baculovirus major envelope glycoprotein
KEYWORDS envelope protein; glycoprotein; transmembrane protein
FEATURE
44,158,189,396 #binding_site carbohydrate (Asn) (covalent) #status predicted
SUMMARY #length 521 #molecular-weight 58675 #checksum 1216
Query Match 81.0%; Score 47; DB 1; Length 521;
Best Local Similarity 66.7%; Pred. No. 4.13e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 40 ERPLNISRI 48
| | | | |
Qy 222 EAPLNVS RV 230

RESULT 2
ENTRY JC4385 #type complete
TITLE LIM protein - rat
ALTERNATE_NAMES CIP-36 protein
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 17-Jan-1996 #sequence_revision 08-Feb-1996 #text_change 10-Oct-1997

ACCESSIONS JC4385
REFERENCE JC4385
#authors Wang, H.; Harrison-Shostak, D.C.; Lemasters, J.J.; Herman, B.
#journal Gene (1995) 165:267-271
#title Cloning of a rat cDNA encoding a novel LIM domain protein

```
#cross-references MUID:96096533
#accession JC4385
#molecule_type mRNA
##residues 1-327 ##label WAN
##cross-references GB:U23769; NID:g1020150; PID:g1020151
##experimental_source hepatocytes
COMMENT This protein is involved in protein-protein interactions and cell
functions, including focal cell adhesion, endocytosis and
maintenance of normal growth control.

GENETICS
#gene clp-36
CLASSIFICATION #superfamily LIM metal-binding repeat homology; GLGF domain
homology
KEYWORDS zinc finger
FEATURE
9-81 #domain GLGF domain homology #label GLG\
258-308 #domain LIM metal-binding repeat homology #label LIM\
258-281 #region zinc finger CCHC motif\
284-308 #region zinc finger CCGH motif
SUMMARY #length 327 #molecular-weight 35525 #checksum 8090

Query Match 79.3%; Score 46; DB 2; Length 327;
Best Local Similarity 66.7%; Pred. No. 6.98e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 25 EOPLAISRV 33
QY 222 EAPLNSRV 230
|||||

RESULT 3
ENTRY A36117 #type complete
TITLE corticosteroid-binding globulin - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change
12-Apr-1996
ACCESSIONS A36117
REFERENCE A36117
#authors Serralini, G.E.; Smith, C.L.; Hammond, G.L.
#journal Mol. Endocrinol. (1990) 4:1166-1172
#title Rabbit corticosteroid-binding globulin: primary structure and
biosynthesis during pregnancy.
#cross-references MUID:91155949
#accession A36117
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-383 ##label SER
CLASSIFICATION #superfamily antithrombin III
SUMMARY #length 383 #molecular-weight 42326 #checksum 3327

Query Match 79.3%; Score 46; DB 2; Length 383;
Best Local Similarity 66.7%; Pred. No. 6.98e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 315 EGPLKVSQV 323
QY 222 EAPLNSRV 230
|||||

RESULT 4
ENTRY E69497 #type complete
TITLE ABC transporter, ATP-binding protein homolog - Archaeoglobus
fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
10-Jul-1998
ACCESSIONS E69497
REFERENCE A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kervlavage, A.R.; Graham,

with high homology to rat RIL.
#cross-references MUID:96096533
#accession JC4385
#molecule_type mRNA
##residues 1-327 ##label WAN
##cross-references GB:U23769; NID:g1020150; PID:g1020151
##experimental_source hepatocytes
COMMENT This protein is involved in protein-protein interactions and cell
functions, including focal cell adhesion, endocytosis and
maintenance of normal growth control.

GENETICS
#gene clp-36
CLASSIFICATION #superfamily LIM metal-binding repeat homology; GLGF domain
homology
KEYWORDS zinc finger
FEATURE
9-81 #domain GLGF domain homology #label GLG\
258-308 #domain LIM metal-binding repeat homology #label LIM\
258-281 #region zinc finger CCHC motif\
284-308 #region zinc finger CCGH motif
SUMMARY #length 327 #molecular-weight 35525 #checksum 8090

Query Match 79.3%; Score 46; DB 2; Length 327;
Best Local Similarity 66.7%; Pred. No. 6.98e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 25 EOPLAISRV 33
QY 222 EAPLNSRV 230
|||||

RESULT 3
ENTRY A36117 #type complete
TITLE corticosteroid-binding globulin - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change
12-Apr-1996
ACCESSIONS A36117
REFERENCE A36117
#authors Serralini, G.E.; Smith, C.L.; Hammond, G.L.
#journal Mol. Endocrinol. (1990) 4:1166-1172
#title Rabbit corticosteroid-binding globulin: primary structure and
biosynthesis during pregnancy.
#cross-references MUID:91155949
#accession A36117
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-383 ##label SER
CLASSIFICATION #superfamily antithrombin III
SUMMARY #length 383 #molecular-weight 42326 #checksum 3327

Query Match 79.3%; Score 46; DB 2; Length 383;
Best Local Similarity 66.7%; Pred. No. 6.98e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 315 EGPLKVSQV 323
QY 222 EAPLNSRV 230
|||||

RESULT 4
ENTRY E69497 #type complete
TITLE ABC transporter, ATP-binding protein homolog - Archaeoglobus
fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
10-Jul-1998
ACCESSIONS E69497
REFERENCE A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kervlavage, A.R.; Graham,

D.E.; Kyripides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession E69497
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-253 ##label KLE
##cross-references GB:AE000966; GB:AE000782; NID:g2689289; PID:g2648557;
TIGR:AF1982
CLASSIFICATION #superfamily ATP-binding cassette homology
P-loop
KEYWORDS
FEATURE
19-210 #domain ATP-binding cassette homology #label ABC\
36-43 #region nucleotide-binding motif A (P-loop)
SUMMARY #length 253 #molecular-weight 27616 #checksum 7436

Query Match 77.6%; Score 45; DB 2; Length 253;
Best Local Similarity 77.8%; Pred. No. 1.17e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 90 EAPLTVSQV 98
QY 222 EAPLNSRV 230
|||||

RESULT 5
ENTRY S71828 #type complete
TITLE gene rll protein - rat
ALTERNATE_NAMES Lim-domain protein rll
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change
16-Dec-1998
ACCESSIONS S71828; S39246
REFERENCE S71828
#authors Schaefer, R.
#submission submitted to the EMBL Data Library, March 1996
#accession S71828
#molecule_type mRNA
#residues 1-330 ##label SCH
#cross-references EMBL:X76454; NID:g755780; PID:g887580
GENETICS
#gene rll
CLASSIFICATION #superfamily LIM metal-binding repeat homology; GLGF domain
homology
FEATURE
9-80 #domain GLGF domain homology #label GLG\
255-305 #domain LIM metal-binding repeat homology #label LIM2
SUMMARY #length 330 #molecular-weight 35521 #checksum 4250

Query Match 77.6%; Score 45; DB 2; Length 330;
Best Local Similarity 75.0%; Pred. No. 1.17e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 25 APLTISRVR 32
QY 223 APLNVSQV 230
|||||

RESULT 6
ENTRY S76909 #type complete
TITLE hypothetical protein - Synecocystis sp. (strain PCC 6803)
```

```

ORGANISM      #formal_name Synchocystis sp.
#variety      PCC 6803
DATE          25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
              21-Aug-1998
ACCESSIONS    S76909
REFERENCE      S74322
#authors      Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
              Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugiura, M.;
              Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
              Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo,
              S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
              Yasuda, M.; Tabata, S.
#journal      DNA Res. (1996) 3:109-136
#title        Sequence analysis of the genome of the unicellular
              cyanobacterium Synchocystis sp. PCC6803. II. Sequence
              determination of the entire genome and assignment of
              potential protein-coding regions.
#cross-references MUID:97061201
#accession     S76909
#status        preliminary
#molecule_type DNA
#residues      1-658 #label KAN
#cross-references EMBL:D90917; GB:AB001339; NID:g1653836; PID:di019554;
              PID:g1653911
#note          the nucleotide sequence was submitted to the EMBL Data
              Library, June 1996
SUMMARY        #length 658 #molecular-weight 75029 #checksum 5182
              77.6%; Score 45; DB 2; Length 658;
              Best Local Similarity 75.0%; Pred. No. 1.17e+01;
              Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 309 DIPLNVS 316
:|||||
Qy 222 EAPLNVS 229

RESULT 7
ENTRY   #type complete
TITLE   ribosomal protein L24 - Bacillus stearothermophilus
ORGANISM #formal_name Bacillus stearothermophilus
DATE     17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change
              30-Sep-1993
ACCESSIONS A02819
REFERENCE   A91149
#authors    Kimura, M.; Kimura, J.; Ashman, K.
#journal     Eur. J. Biochem. (1985) 150:491-497
#title       The complete primary structure of ribosomal proteins L1, L14,
              L15, L23, L24, and L29 from Bacillus stearothermophilus.
#cross-references MUID:85257681
#accession     A02819
#molecule_type protein
#residues      1-103 #label KIM
CLASSIFICATION #superfamily Escherichia coli ribosomal protein L24
KEYWORDS       protein biosynthesis; ribosome
SUMMARY        #length 103 #molecular-weight 11218 #checksum 7478
              75.9%; Score 44; DB 1; Length 103;
              Best Local Similarity 66.7%; Pred. No. 1.94e+01;
              Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 60 EAPIHVS 68
:|||||
Qy 222 EAPLNVS 230

RESULT 8
ENTRY   #type complete
TITLE   MyD88 protein - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE     18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
              10-Sep-1997
ACCESSIONS S11226; JC2079

```

```

REFERENCE      S11226
#authors      Lord, K.A.; Hoffman-Liebermann, B.; Liebermann, D.A.
#journal       Oncogene (1990) 5:1095-1097
#title         Nucleotide sequence and expression of a cDNA encoding MyD88,
              a novel myeloid differentiation primary response gene
              induced by IL6.
#cross-references MUID:90326414
#accession     S11226
#molecule_type mRNA
#residues      1-243 #label LOR
#cross-references EMBL:X51397; NID:g53293; PID:g53294
REFERENCE      JC2079
#authors      Hultmark, D.
#journal       Biochem. Biophys. Res. Commun. (1994) 199:144-146
#title         Macrophage differentiation marker MyD88 is a member of the
              Toll/IL-1 receptor family.
#cross-references MUID:94168566
#contents      annotation
SUMMARY        #length 243 #molecular-weight 28039 #checksum 4521
              75.9%; Score 44; DB 2; Length 243;
              Best Local Similarity 66.7%; Pred. No. 1.94e+01;
              Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 73 EKPQVAR 81
:|||||
Qy 222 EAPLNVS 230

RESULT 9
ENTRY   #type complete
TITLE   fimbriin precursor - Porphyromonas gingivalis
ORGANISM #formal_name Porphyromonas gingivalis
DATE     14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change
              09-Sep-1997
ACCESSIONS JN0918; B60275
REFERENCE   JN0913
#authors    Fujiwara, T.; Morishima, S.; Takahashi, I.; Hamada, S.
#journal     Biochem. Biophys. Res. Commun. (1993) 197:241-247
#title       Molecular cloning and sequencing of the fimbriin gene of
              Porphyromonas gingivalis strains and characterization of
              recombinant proteins.
#cross-references MUID:94071950
#accession     JN0918
#molecule_type DNA
#residues      1-348 #label FUJ
#cross-references DBJ:D17799; NID:g456500; PID:di005154; PID:g456501
#experimental_source strain OMZ409
REFERENCE      A60275
#authors      Lee, J.Y.; Sojar, H.T.; Bedi, G.S.; Genco, R.J.
#journal       Infect. Immun. (1991) 59:383-389
#title         Porphyromonas (Bacteroides) gingivalis fimbriin: size,
              amino-terminal sequence, and antigenic heterogeneity.
#cross-references MUID:91099990
#accession     B60275
#status        preliminary
#molecule_type protein
#residues      11-30 #label LEE
#experimental_source strain A7A1-28
GENETICS
#gene          fimA
#start_codon   GTG
FUNCTION        #description this protein mediates cell adhesion to oral tissues and is
              important in colonization
              cell adhesion
KEYWORDS        #domain signal sequence #status predicted #label SIG\
              1-9
              #product fimbriin #status predicted #label MAT
              10-348
              #length 348 #molecular-weight 38076 #checksum 6153
SUMMARY

Query Match 75.9%; Score 44; DB 2; Length 348;
Best Local Similarity 44.4%; Pred. No. 1.94e+01;

```

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 124 DTPLEIKRV 132
:||||: ||
QY 222 EAPLNVSRV 230

RESULT 10
ENTRY #type complete
TITLE fimbrilin precursor - Porphyromonas gingivalis
ORGANISM #formal_name Porphyromonas gingivalis
DATE 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 09-Sep-1997
JN0916; JN0917
ACCESSIONS JN0916; JN0917
REFERENCE Fujiwara, T.; Morishima, S.; Takahashi, I.; Hamada, S.
#authors Biochem. Biophys. Res. Commun. (1993) 197:241-247
#journal Molecular cloning and sequencing of the fimbrilin gene of
#title Porphyromonas gingivalis strains and characterization of recombinant proteins.
#cross-references MUID:94071950
#accession JN0916
#molecule_type DNA
#residues 1-348 #label FUJ
#cross-references DBJ:DL17797; NID:g456497; PID:dl005152; PID:g509036
#experimental_source strain HW24D1
#accession JN0917
#molecule_type DNA
#residues 1-347 #label FU2
#cross-references DBJ:DL17798; NID:g456498; PID:dl005153; PID:g456499
#experimental_source strain OMZ314

GENETICS
#gene fimA
#start_codon GTG

FUNCTION
#description this protein mediates cell adhesion to oral tissues and is important in colonization

KEYWORDS
FEATURE 1-9 #domain signal sequence #status predicted #label SIG\
10-348 #product fimbrilin #status predicted #label MAT\
10-347 #product fimbrilin #status predicted #label MA2
SUMMARY #length 348 #molecular-weight 38089 #checksum 5420

Query Match 75.9%; Score 44; DB 2; Length 348;
Best Local Similarity 44.4%; Pred. No. 1.94e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 124 DTPLEIKRV 132
:||||: ||
QY 222 EAPLNVSRV 230

RESULT 11
ENTRY #type complete
TITLE fimbrilin precursor - Porphyromonas gingivalis (ATCC 49417)
ORGANISM #formal_name Porphyromonas gingivalis
DATE 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 09-Sep-1997
JN0919
ACCESSIONS JN0919
REFERENCE Fujiwara, T.; Morishima, S.; Takahashi, I.; Hamada, S.
#authors Biochem. Biophys. Res. Commun. (1993) 197:241-247
#journal Molecular cloning and sequencing of the fimbrilin gene of
#title Porphyromonas gingivalis strains and characterization of recombinant proteins.
#cross-references MUID:94071950
#accession JN0919
#molecule_type DNA
#residues 1-350 #label FUJ
#cross-references DBJ:DL17800; NID:g456502; PID:dl005155; PID:g456503
#experimental_source ATCC 49417

COMMENT This protein mediates cell adhesion to oral tissues and is

important in colonization.

GENETICS
#gene fimA
#start_codon GTG
KEYWORDS cell adhesion
FEATURE 1-9
SUMMARY #domain signal sequence #status predicted #label SIG\
10-350 #product fimbrilin #status predicted #label MAT
#length 350 #molecular-weight 37911 #checksum 2707

Query Match 75.9%; Score 44; DB 2; Length 350;
Best Local Similarity 44.4%; Pred. No. 1.94e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 126 DTPLEIKRV 134
:||||: ||
QY 222 EAPLNVSRV 230

RESULT 12
ENTRY #type complete
TITLE hypothetical protein RV2030c - Mycobacterium tuberculosis (strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
ACCESSIONS E70942
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
#cross-references MUID:98295987
#accession E70942
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-681 #label COL
##cross-references GB:AL021899; GB:AL123456; NID:g3242282; PID:el252009; PID:g2896767
##experimental_source strain H37RV

GENETICS
#gene RV2030c
SUMMARY #length 681 #molecular-weight 74930 #checksum 3008

Query Match 75.9%; Score 44; DB 2; Length 681;
Best Local Similarity 66.7%; Pred. No. 1.94e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 612 EAPLDVRL 620
||||: ||
QY 222 EAPLNVSRV 230

RESULT 13
ENTRY #type complete
TITLE hypothetical protein PH0162 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
H71237
ACCESSIONS H71237
REFERENCE A71000
#authors Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;


```

Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. (1998) 5:55-76
#journal
#title
Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references MUID:98344137
#accession H71237
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-690 #label KAW
#cross-references GB:AP000001; NID:g3236128; PID:d1030174; PID:g3256548
#experimental_source strain OT3
#note this accession replaces an interim accession for a
sequence replaced by GenBank
GENETICS
#gene PH0162
SUMMARY
#length 690 #molecular-weight 79695 #checksum 4137
Query Match 75.9%; Score 44; DB 2; Length 690;
Best Local Similarity 44.4%; Pred. No. 1.94e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 92 ETPKIDRI 100
I::: I:
Qy 222 EAPLNSRV 230

RESULT 14
ENTRY
TITLE
ORGANISM
DATE
13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
08-Sep-1997
ACCESSIONS
#gene S43192; S15438
#accession S43192
#status preliminary
#molecule_type DNA
#residues 1-110 #label VVS
#cross-references EMBL:Z31726; NID:g459675; PID:g459676
REFERENCE
#authors Jahn, O.; Hartmann, R.K.; Erdmann, V.A.
#journal Eur. J. Biochem. (1991) 197:733-740
#title Analysis of the spc ribosomal protein operon of Thermus
aquaticus.
#cross-references MUID:91231015
#accession S15438
#molecule_type DNA
#residues 1, 'QA', 4, 'V', 6-28, 'A', 30-32, 'R', 34, 'M', 36-43, 'L', 45-50,
'P', 52-54, 'H', 56-60, 'V', 62, 'Q', 64-91, 'D', 93, 'R', 95-97,
'A', 99-104, 'S', 106-107, 'V', 109-110 #label JAH
#cross-references EMBL:X56552; NID:g48102; PID:g48105
CLASSIFICATION #superfamily Escherichia coli ribosomal protein L24
KEYWORDS protein biosynthesis; ribosome
SUMMARY
#length 110 #molecular-weight 12056 #checksum 5830
Query Match 74.1%; Score 43; DB 2; Length 110;
Best Local Similarity 66.7%; Pred. No. 3.20e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 64 EAPLHASKV 72
I::: I:
Qy 222 EAPLNSRV 230

```

```

RESULT 15
ENTRY
TITLE
ORGANISM
DATE
04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
21-Nov-1998
ACCESSIONS
#gene E42645
#accession E42645
#status preliminary
#molecule_type DNA
#residues 1-111 #label KAU
#cross-references GB:M80325; NID:g144617; PID:g144622
#note sequence extracted from NCBI backbone (NCBIN:79464,
NCBIP:79469)
REFERENCE
#authors Stephens, R.S.; Kallman, S.; Lammel, C.J.; Fan, J.; Marathe,
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:734-739
#title Genome sequence of an obligate intracellular pathogen of
humans: Chlamydia trachomatis.
#cross-references MUID:99000809
#accession C71506
#molecule_type DNA
#residues 1-84, 'R', 86-104, 'S', 106-111 #label ARN
#cross-references GB:AE001323; GB:AE001273; NID:g3328931; PID:g3328954
#experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene rl24
CLASSIFICATION #superfamily Escherichia coli ribosomal protein L24
KEYWORDS protein biosynthesis; ribosome
SUMMARY
#length 111 #molecular-weight 12608 #checksum 3205
Query Match 74.1%; Score 43; DB 2; Length 111;
Best Local Similarity 66.7%; Pred. No. 3.20e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 61 EAPLHISNV 69
I::: I:
Qy 222 EAPLNSRV 230
Search completed: Thu Jul 8 18:34:19 1999
Job time : 8 secs.

```

THIS PAGE BLANK (USPTO)

WQSEKHA

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:33:14 1999; Maspar time 3.37 Seconds
Tabular output not generated.
75.550 Million cell updates/sec

Title: >US-09-041-236-2
Description: (222-230) from US09041236.pep (14 of 45)
Perfect Score: 58
Sequence: 1 EAPLNVSrv 9

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28269293 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 22.600; Variance 21.385; scale 1.057

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	47	81.0	521	1	VENV_DHV11	1.37e+00	
2	46	79.3	327	1	CL36_RAT	2.45e+00	
3	46	79.3	383	1	CBG_RABIT	2.45e+00	
4	45	77.6	167	1	THIM_MAIZE	4.34e+00	
5	45	77.6	328	1	RIL_HUMAN	4.34e+00	
6	45	77.6	330	1	RIL_RAT	4.34e+00	
7	45	77.6	525	1	GOR_PANTR	4.34e+00	
8	44	75.9	103	1	RL24_BAGST	7.60e+00	
9	44	75.9	270	1	PRC2_ORYSA	7.60e+00	
10	44	75.9	296	1	MY88_MOUSE	7.60e+00	
11	43	74.1	110	1	RL24_THETH	1.32e+01	
12	43	74.1	111	1	FLAV_ENTAG	1.32e+01	
13	43	74.1	177	1	FLAV_ENTAG	1.32e+01	
14	43	74.1	177	1	FLAV_ENTAG	1.32e+01	
15	43	74.1	329	1	MKK2_CRILLO	1.32e+01	
16	43	74.1	362	1	SERC_SALTY	1.32e+01	
17	43	74.1	362	1	PHOSPHOSERINE AMINOTRA	1.32e+01	
18	43	74.1	385	1	MKK2_MOUSE	1.32e+01	
19	43	74.1	391	1	HTPG_VIBFI	1.32e+01	
20	43	74.1	400	1	MKK2_HUMAN	1.32e+01	
21	43	74.1	400	1	ENPL_MESNU	1.32e+01	
22	43	74.1	518	1	ATPA_ENTHR	1.32e+01	
23	43	74.1	616	1	HTPG_BORBU	1.32e+01	

24	43	74.1	621	1	HTPG_HELPY	HEAT SHOCK PROTEIN HTP	1.32e+01
25	43	74.1	624	1	HTPG_ECOLI	HEAT SHOCK PROTEIN HTP	1.32e+01
26	43	74.1	626	1	HTPG_ACTAC	HEAT SHOCK PROTEIN HTP	1.32e+01
27	43	74.1	631	1	HTPG_HAEN	HEAT SHOCK PROTEIN HTP	1.32e+01
28	43	74.1	639	1	HTPG_TREPA	HEAT SHOCK PROTEIN HTP	1.32e+01
29	43	74.1	716	1	ENPL_RABIT	ENDOPLASMIN (94 KD GLU	1.32e+01
30	43	74.1	795	1	ENPL_CHICK	ENDOPLASMIN PRECURSOR	1.32e+01
31	43	74.1	802	1	ENPL_MOUSE	ENDOPLASMIN PRECURSOR	1.32e+01
32	43	74.1	803	1	ENPL_HUMAN	ENDOPLASMIN PRECURSOR	1.32e+01
33	43	74.1	804	1	ENPL_PIG	ENDOPLASMIN PRECURSOR	1.32e+01
34	43	74.1	804	1	ENPL_CANFA	ENDOPLASMIN PRECURSOR	1.32e+01
35	43	74.1	1530	1	RPCL_TRYBB	DNA-DIRECTED RNA POLYM	1.32e+01
36	43	74.1	4725	1	DIHC_DICDI	DYNEIN HEAVY CHAIN, CY	1.32e+01
37	42	72.4	234	1	YIIM_ECOLI	HYPOTHETICAL 26.6 KD P	2.26e+01
38	42	72.4	255	1	ETFB_HUMAN	ELECTRON TRANSFER FLAV	2.26e+01
39	42	72.4	441	1	YXEK_BACSU	HYPOTHETICAL 49.3 KD P	2.26e+01
40	42	72.4	781	1	TORB_BREDI	ISOQUINOLINE 1-OXIDORE	2.26e+01
41	42	72.4	795	1	Y210_HUMAN	HYPOTHETICAL PROTEIN K	2.26e+01
42	41	70.7	332	1	YI36_MYCTU	HYPOTHETICAL 35.3 KD P	3.83e+01
43	41	70.7	699	1	HS82_ORYSA	HEAT SHOCK PROTEIN 82	3.83e+01
44	41	70.7	702	1	YAI4_YEAST	HYPOTHETICAL 78.3 KD P	3.83e+01
45	41	70.7	721	1	HS90_THEPA	HEAT SHOCK PROTEIN 90	3.83e+01

ALIGNMENTS

RESULT 1
ID VENV_DHV11 STANDARD; PRT; 521 AA.
AC P27427;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DE 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN PRECURSOR.
OS DHORI VIRUS (STRAIN INDIAN/1313/61) (DHO).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; ORTHOMYXOVIRIDAE;
OC THOGOTO-LIKE VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX FREEDMAN-FALUTSCH E.Z., FULLER F.J.;
RT "Nucleotide sequence of the tick-borne, orthomyxo-like
Dhori/Indian/1313/61 virus envelope gene."
RL Virology 175:10-18(1990).
CC -!- FUNCTION: POSSIBLE ROLE IN ENDOCYTOTIC FUSION EVENTS DURING
INFECTION.
CC -!- SUBUNIT: DISULFIDE-LINKED HOMODIMER (POSSIBLY HOMODIMER).
CC -!- SIMILARITY: TO THOGOTO VIRUS ENVELOPE GLYCOPROTEIN AND TO
BACULOVIRUSES MAJOR ENVELOPE GLYCOPROTEIN (P64/P67).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M34002; G335190; -
DR PIR: A34679; VGI0VH.
DR GLYCOPROTEIN; TRANSMEMBRANE; SIGNAL.
KW SIGNAL.
FT CHAIN 1 20 POTENTIAL.
FT SIGNAL 21 521 ENVELOPE GLYCOPROTEIN.
FT TRANSME 501 517 POTENTIAL.
FT CARBOHYD 44 44 POTENTIAL.
FT CARBOHYD 158 158 POTENTIAL.
FT CARBOHYD 189 189 POTENTIAL.
FT CARBOHYD 396 396 POTENTIAL.
SQ SEQUENCE 521 AA; 58676 MW; B94E2693 CRC32;

Query Match 81.0%; Score 47; DB 1; Length 521;
Best Local Similarity 66.7%; Pred. No. 1.37e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Db 40 EKPLNISRI 48
QY 222 EAPLNSRV 230

RESULT 2
ID CL36_RAT STANDARD; PRT; 327 AA.
AC P52944;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE LIM PROTEIN CLP36.
GN CLP36.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN: SPRAGUE-DAWLEY; TISSUE=LIVER;
RX MEDLINE; 96096533.
RA WANG H., HARRISON-SHOSHTAK D.C., LEMASTERS J.J., HERMAN B.;
RT "Cloning of a rat cDNA encoding a novel LIM domain protein with high
homology to rat RIL".
RL GENE 165:267-271(1995).
CC -!- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN HEART, LUNG AND
LIVER, MODERATELY IN SPLEEN AND SKELETAL MUSCLE, AND AT EXTREMELY
LOW LEVELS (IF AT ALL) IN TESTIS AND BRAIN TISSUES.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U23769; G1020151; -
DR PROSITE; PS00478; LIM_DOMAIN_1; FALSE_NEG.
DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
DR PFAM; PF00412; LIM; 1.
DR PFAM; PF00595; PD2; 1.
KW LIM MOTIF; METAL-BINDING; ZINC.
FT DOMAIN 256 306 LIM.
FT SEQUENCE 327 AA; 35525 MW; D11421B1 CRC32;
SQ SEQUENCE 327 AA; 35525 MW; D11421B1 CRC32;

Query Match 79.3%; Score 46; DB 1; Length 327;
Best Local Similarity 66.7%; Pred. No. 2.45e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 25 EOPLAISRV 33
QY 222 EAPLNSRV 230

RESULT 3
ID CBG_RABIT STANDARD; PRT; 383 AA.
AC P23775;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE CORTICOSTEROID-BINDING GLOBULIN (CBG) (TRANSCORTIN).
GN CBG.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE OF 15-383 FROM N.A., AND SEQUENCE OF 1-24.
RC TISSUE=LIVER;
RX MEDLINE; 91155949.
RA SERALINI G.-E., SMITH C.L., HAMMOND G.L.;
RT "Rabbit corticosteroid-binding globulin: primary structure and
```

```
RT biosynthesis during pregnancy.";
RL MOL. ENDOCRINOL. 4:1166-1172(1990).
CC -!- FUNCTION: MAJOR TRANSPORT PROTEIN FOR GLUCOCORTICOIDS AND
PROGESTINS IN THE BLOOD OF ALMOST ALL VERTEBRATE SPECIES.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: PRODUCED AND SECRETED BY HEPATOCYTES, BUT HAS
ALSO BEEN IDENTIFIED IN A NUMBER OF GLUCOCORTICOID RESPONSIVE
CELLS (IT IS FOUND IN MATERNAL LUNG, SPLEEN, AND OVARY AND FETAL
KIDNEY).
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR PIR; A36117; A36117.
DR PROSITE; PS00284; SERPIN; 1.
DR PFAM; PF00079; serpin; 1.
DR HSP; P05154; IPA1.
KW STEROID-BINDING; TRANSPORT; PLASMA; SERPIN; GLYCOPROTEIN.
FT CARBOHYD 74 74 POTENTIAL.
FT CARBOHYD 154 154 POTENTIAL.
FT CARBOHYD 238 238 POTENTIAL.
FT CARBOHYD 308 308 POTENTIAL.
FT SITE 228 228 CONSERVED CYSTEINE WITHIN STEROID
BINDING DOMAIN (BY SIMILARITY).
SQ SEQUENCE 383 AA; 42326 MW; 43F330C8 CRC32;

Query Match 79.3%; Score 46; DB 1; Length 383;
Best Local Similarity 66.7%; Pred. No. 2.45e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 315 EGPLKSVK 323
QY 222 EAPLNSRV 230

RESULT 4
ID THIM_MAIZE STANDARD; PRT; 167 AA.
AC Q41864;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE THIREDOXIN M-TYPE PRECURSOR (TRX-M).
GN TRM1.
OS ZEA MAYS (MAIZE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; ZEA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. B23; TISSUE=LEAF;
RA TREYANON S.J., ASHTON A.R.;
RT "Isolation of a full-length cDNA clone for thioredoxin-m from maize.";
RL (IN) PLANT GENE REGISTER PGR95-129.
CC -!- FUNCTION: THIREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS.
THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -!- THERE ARE THREE TYPES OF THIREDOXINS IN PLANTS: THE H-TYPE WHICH
IS CYTOSOLIC AND INVOLVED IN HETEROGRAPHIC PROCESSES, THE M-TYPE
WHICH ACTIVATES NADP-MALATE DEHYDROGENASE AND THE F-TYPE WHICH
ACTIVATES ENZYMES OF THE PHOTOSYNTHETIC CARBON CYCLE. BOTH THE M-
AND F-TYPES ARE LOCATED IN THE CHLOROPLAST.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L40957; G725276; -
DR MAIZEDB; 114054; -
DR PROSITE; PS00194; THIREDOXIN; 1.
DR PFAM; PF00085; thired; 1.
DR HSP; P00274; 1XOA.
```

```

KW REDOX-ACTIVE CENTER; ELECTRON TRANSPORT; CHLOROPLAST; TRANSIT PEPTIDE.
FT TRANSIT 1 53 CHLOROPLAST (POTENTIAL).
FT CHAIN 54 167 THIOREDOXIN M-TYPE.
FT DISULFID 89 92 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 167 AA; 18073 MW; 9B71143E CRC32;

Query Match 77.6%; Score 45; DB 1; Length 167;
Best Local Similarity 75.0%; Pred. No. 4.34e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 36 APLSVGRV 43
   |||:|:|
QY 223 APLNSRV 230

RESULT 5
ID RLL_HUMAN STANDARD; PRT; 328 AA.
AC P50479;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE LIM PROTEIN RIL.
GN RIL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA SCHARM B., SCHAEFER R.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 1-20 FROM N.A.
RA BASHIROVA A.A., MARKELOV M.L., SHLIKOVA T.V., LEVSHENKOVA E.V.,
RA FROLOVA E.I.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X93510; E211802; -
CC DR EMBL; U82997; G1786093; -
CC DR PROSITE; P500478; LIM_DOMAIN_1; 1.
CC DR PROSITE; P500023; LIM_DOMAIN_2; 1.
CC DR PFAM; PF00412; LIM; 1.
CC DR PFAM; PF00595; PDZ; 1.
CC DR HSP; Q12959; 1PDR.
KW LIM MOTIF; METAL-BINDING; ZINC.
FT DOMAIN 253 303 LIM.
SQ SEQUENCE 328 AA; 35082 MW; B39DA148 CRC32;

Query Match 77.6%; Score 45; DB 1; Length 328;
Best Local Similarity 75.0%; Pred. No. 4.34e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 24 APLTISRV 31
   |||:|:|
QY 223 APLNSRV 230

RESULT 6
ID RIL_RAT STANDARD; PRT; 330 AA.
AC P36202;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
CC
```

```

DE LIM PROTEIN RIL (RIT-18) (H-REV18).
GN RIL.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RX MEDLINE; 95124724.
RA KIESS M., SCHARM B., AGUZZI A., HAJNAL A., KLEMENZ R.,
RA SCHWARTE-WALDHOFF I., SCHAEFER R.;
RT "Expression of ril, a novel LIM domain gene, is down-regulated in
RT Hras-transformed cells and restored in phenotypic revertants.";
RL ONCOGENE 10:61-68(1995).
CC -!- TISSUE SPECIFICITY: DETECTED IN SEVERAL TISSUES, MOST PROMINENT IN
CC BRAIN AND HEART OF ADULTS.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X76454; G887580; -
CC DR PIR; S39246; S39246.
CC DR PROSITE; P500478; LIM_DOMAIN_1; 1.
CC DR PROSITE; P500023; LIM_DOMAIN_2; 1.
CC DR PFAM; PF00412; LIM; 1.
CC DR PFAM; PF00595; PDZ; 1.
KW LIM MOTIF; METAL-BINDING; ZINC.
FT DOMAIN 255 305 LIM.
SQ SEQUENCE 330 AA; 35521 MW; 6A4E4D01 CRC32;

Query Match 77.6%; Score 45; DB 1; Length 330;
Best Local Similarity 75.0%; Pred. No. 4.34e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 25 APLTISRV 32
   |||:|:|
QY 223 APLNSRV 230

RESULT 7
ID GOR_PANTR STANDARD; PRT; 525 AA.
AC P48778;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ANTIGEN GOR (FRAGMENT).
GN GOR.
OS PAN TROGLODYTES (CHIMPANZEE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; PAN.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91061544.
RA MISHIRO S., HOSHI Y., TAKEDA K., YOSHIKAWA A., GOTANDA T.,
RA TAKAHASHI K., AKAHANE Y., YOSHIZAWA H., OKAMOTO H., TSUDA F.,
RA PETERSON D.A., MUCHMORE E.;
RT "Non-A, non-B hepatitis specific antibodies directed at host-derived
RT epitope: implication for an autoimmune process.";
RL LANCET 336:1400-1403(1990).
RN [2]
RP ERRATUM.
RA MISHIRO S., HOSHI Y., TAKEDA K., YOSHIKAWA A., GOTANDA T.,
RA TAKAHASHI K., AKAHANE Y., YOSHIZAWA H., OKAMOTO H., TSUDA F.,
RA PETERSON D.A., MUCHMORE E.;
RL LANCET 337:252-252(1990).
CC -!- DISEASE: INDIVIDUALS WITH HEPATITIS C (NON-A, NON-B HEPATITIS;
```

CC NANBH) HAVE ANTIBODIES AGAINST GOR.
 CC -!- SIMILARITY: TO C.ELEGANS PROTEIN F52C9.4 AND YEAST YGR276C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D1001370; -.
 DR PFAM: PF00929; Exonuclease; 1.
 KW ANTIGEN.
 FT NON_TER 1
 FT SITE 447 461 GOR EPIPOPE.
 SQ SEQUENCE 525 AA; 57712 MW; CFALDAC7 CRC32;

 Query Match 77.6%; Score 45; DB 1; Length 525;
 Best Local Similarity 66.7%; Pred. No. 4.34e+00;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 Db 495 EPPLPVRV 503
 QY 222 EAPLNVSRV 230

 RESULT 8
 ID RL24_BACST STANDARD; PRT; 103 AA.
 AC P04455;
 DT 13-AUG-1987 (REL. 05, CREATED)
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
 DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
 DE 50S RIBOSOMAL PROTEIN L24.
 GN RPLX.
 OS BACILLUS STEAROTHERMOPHILUS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 OC BACILLUS.
 RN [1]
 RP SEQUENCE.
 RX MEDLINE; 85257681.
 RA KIMURA M., KIMURA J., ASHMAN K.;
 RT "The complete primary structure of ribosomal proteins L1, L14, L15,
 RT L23, L24 and L29 from *Bacillus stearothermophilus*.";
 RL EUR. J. BIOCHEM. 150:491-497(1985).
 CC -!- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
 CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
 CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT.
 CC -!- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
 DR PIR; A02819; R5BS24.
 DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
 DR PFAM; PF00467; L24; 1.
 KW RIBOSOMAL PROTEIN.
 SQ SEQUENCE 103 AA; 11218 MW; 1DCC7387 CRC32;

 Query Match 75.9%; Score 44; DB 1; Length 103;
 Best Local Similarity 66.7%; Pred. No. 7.60e+00;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

 Db 60 EAPIHVKV 68
 QY 222 EAPLNVSRV 230

 RESULT 9
 ID PRC2_ORYSA STANDARD; PRT; 270 AA.
 AC P52428;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE PROTEASOME COMPONENT C2 (EC 3.4.99.46) (MACROPAIN SUBUNIT C2)
 DE (MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT C2).
 OS ORYZA SATIVA (RICE).

OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
 OC POACEAE; ORYZA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA UMEDA M., FUJII N., UCHIMIYA H.;
 RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
 CC WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH
 CC ARG. PHE. TYR. LEU. AND GLU ADJACENT TO THE LEAVING GROUP AT
 CC NEUTRAL OR SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT
 CC PROTEOLYTIC ACTIVITY.
 CC -!- PATHWAY: IS INVOLVED IN AN ATP/UBIQUITIN-DEPENDENT NON-LYSOSOMAL
 CC PROTEOLYTIC PATHWAY.
 CC -!- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL
 CC SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE.
 CC -!- SUBCELLULAR LOCATION: PROTEASOMES ARE FOUND IN THE CYTOPLASM AND
 CC ALSO IN THE NUCLEUS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A; ALSO KNOWN AS THE
 CC PROTEASOME A-TYPE FAMILY. BELONGS TO THE C2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D37886; D1007705; -.
 DR PROSITE; PS00388; PROTEASOME_A; 1.
 DR PFAM; PF00227; Proteasome; 1.
 DR HSP; P25156; 1PWA.
 KW PROTEASOME; HYDROLASE; PROTEASE.
 SQ SEQUENCE 270 AA; 29630 MW; D227F9BA CRC32;

 Query Match 75.9%; Score 44; DB 1; Length 270;
 Best Local Similarity 66.7%; Pred. No. 7.60e+00;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 Db 100 DAPLPVSR 108
 QY 222 EAPLNVSRV 230

 RESULT 10
 ID MY88_MOUSE STANDARD; PRT; 296 AA.
 AC P22366;
 DT 01-AUG-1991 (REL. 19, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE MYELOID DIFFERENTIATION PRIMARY RESPONSE PROTEIN MYD88.
 GN MYD88.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97165970.
 RA BONNET T.P., GARKA K.E., PARNET P., SONODA G., TESTA J.R.,
 RA SIMS J.E.;
 RT "The cloning and characterization of human MyD88: a member of an IL-1
 RT receptor related family";
 RL FEBS LETT. 402:81-81(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA BONNET T.P., GARKA K.E., PARNET P., HOLLINGSWORTH T., HUANG C.P.,
 RA TIMOUR M., SIMS J.E.;
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE OF 17-296 FROM N.A.
 RX MEDLINE; 90326414.
 RA LORD K.A., HOFFMAN-LIEBERMANN B., LIEBERMANN D.A.;

Nucleotide sequence and expression of a cDNA encoding MyD88, a novel myeloid differentiation primary response gene induced by IL6.;
ONCOGENE 5:1095-1097(1990).
CC -!- TISSUE SPECIFICITY: MRELOID PRECURSOR ENRICHED MURINE BONE-MARROW.
CC -!- INDUCTION: BY INTERLEUKIN-6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U84409; G1814022; -.
DR EMBL; X51397; G53294; -.
DR PIR; S11226; S11226.
DR MGD; MGI:108005; MYD88.
DR PROSITE; PS50017; DEATH_DOMAIN; 1.
DR PFAM; PF00531; death; 1.
DR HSP; P07174; INGR.
KW DIFFERENTIATION.
FT DOMAIN 54 109 DEATH DOMAIN.
SQ SEQUENCE 296 AA; 33779 MW; E0F5717F CRC32;

Query Match 75.9%; Score 44; DB 1; Length 296;
Best Local Similarity 66.7%; Pred. No. 7.60e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 126 EKPQVARV 134
| | | | |
QY 222 EAPLNSRV 230

RESULT 11
ID RL24_THETH STANDARD; PRT; 110 AA.
AC P24318;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE 50S RIBOSOMAL PROTEIN L24.
GN RPLX OR RPL24.
OS THERMUS AQUATICUS (SUBSP. THERMOPHILUS).
OC BACTERIA; THERMUS/DEINOCOCCUS GROUP; THERMUS GROUP; THERMUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB8;
RX MEDLINE; 91231015.
RA JAHN O., HARTMANN R.K., ERDMANN V.A.;
RT "Analysis of the spc ribosomal protein operon of Thermus aquaticus.";
RL EUR. J. BIOCHEM. 197;733-740(1991).
RN [2]
RP SEQUENCE OF 1-44.
RC STRAIN=HB8;
RA BOYSEN R.I., SCHROEDER W., ERDMANN V.A.;
RL SUBMITTED (FEB-1994) TO THE SWISS-PROT DATA BANK.
CC -!- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56552; G48105; -.
DR PIR; S15438; S15438.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.

DR PFAM; PF00467; L24; 1.
KW RIBOSOMAL PROTEIN.
FT CONFLICT 2 3 QA -> RV (IN REF. 2).
FT CONFLICT 5 5 V -> M (IN REF. 2).
FT CONFLICT 29 29 A -> E (IN REF. 2).
FT CONFLICT 33 33 R -> K (IN REF. 2).
FT CONFLICT 35 35 M -> Y (IN REF. 2).
FT CONFLICT 44 44 L -> I (IN REF. 2).
SQ SEQUENCE 110 AA; 11879 MW; 0B1CA5C CRC32;

Query Match 74.1%; Score 43; DB 1; Length 110;
Best Local Similarity 66.7%; Pred. No. 1.32e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 64 EAPLHASKV 72
| | | | |
QY 222 EAPLNSRV 230

RESULT 12
ID RL24_CHLTR STANDARD; PRT; 111 AA.
AC P28537;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 50S RIBOSOMAL PROTEIN L24.
GN RPLX.
OS CHLAMYDIA TRACHOMATIS.
OC BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=434/BU / SEROVAR L2;
RX MEDLINE; 92138612.
RA KAUL R., GRAY G.J., KOENCKE N.R., GU L.J.;
RT "Cloning and sequence analysis of the Chlamydia trachomatis spc
RT ribosomal protein gene cluster".
RL J. BACTERIOL. 174:1205-1212(1992).
CC -!- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80325; G144622; -.
DR PIR; E42645; E42645.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
DR PFAM; PF00487; L24; 1.
KW RIBOSOMAL PROTEIN.
SQ SEQUENCE 111 AA; 12608 MW; 0D7B342A CRC32;

Query Match 74.1%; Score 43; DB 1; Length 111;
Best Local Similarity 66.7%; Pred. No. 1.32e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 61 EAPLHISNV 69
| | | | |
QY 222 EAPLNSRV 230

RESULT 13
ID FLAV_ENTAG STANDARD; PRT; 177 AA.
AC P28579;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE FLAVODOXIN.
GN NIFF.
OS ENTEROBACKER AGGLOMERANS.
OG PLASMID PEAS.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ENTEROBACKER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-333;
RX MEDLINE: 91217003.
RA KREUTZER R., DAYANANDA S., KLINGMUELLER W.;
RT "Cloning of the electron transport protein genes nifJ and
RL nifH in Enterobacter agglomerans 333.";
RL J. BACTERIOL. 173:3252-3256(1991).
RN [2]
RP SEQUENCE OF 1-21 FROM N.A.
RC STRAIN-333;
RA SCHWICKERATH O.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: LOW-POTENTIAL ELECTRON DONOR TO A NUMBER OF REDOX
CC ENZYMES. NIFF IS THE ELECTRON DONOR TO NITROGENASE.
CC -!- COFACTOR: FMN.
CC -!- SIMILARITY: BELONGS TO THE FLAVODOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DE EMBL: M38221; G145134; -
DR EMBL: X96964; E256968; -
DR EMBL: X78558; G469076; -
DR PIR: A39414; A39414.
DR PROSITE: PS00201; FLAVODOXIN: 1.
DR PFAM: PF00258; flavodoxin: 1.
DR HSSP: P10340; 10FV.
KW NITROGEN FIXATION; ELECTRON TRANSPORT; FLAVOPROTEIN; FMN; PLASMID.
SQ SEQUENCE 177 AA; 19581 MW; DA51F3D3 CRC32;
Query Match 74.1%; Score 43; DB 1; Length 177;
Best Local Similarity 62.5%; Pred. No. 1.32e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 31 DAPLDVRR 38
QY :|||:|
QY 222 EAPLNVS 229
RESULT 14
ID FLAV_ENTAG STANDARD; PRT: 177 AA.
AC P71169;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE FLAVODOXIN.
GN NIFF.
OS ENTEROBACKER AGGLOMERANS.
OG PLASMID PEAS.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ENTEROBACKER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-333;
RX STEIBL H.D., SIDDAVATTAM D.;
RA SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RL -!- FUNCTION: LOW-POTENTIAL ELECTRON DONOR TO A NUMBER OF REDOX
CC ENZYMES. NIFF IS THE ELECTRON DONOR TO NITROGENASE.
CC -!- COFACTOR: FMN.
CC -!- SIMILARITY: BELONGS TO THE FLAVODOXIN FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DE EMBL: X81894; E259010; -
DR PROSITE: PS00201; FLAVODOXIN: 1.
DR PFAM: PF00258; flavodoxin: 1.
KW NITROGEN FIXATION; ELECTRON TRANSPORT; FLAVOPROTEIN; FMN; PLASMID.
SQ SEQUENCE 177 AA; 19584 MW; 9A99A991 CRC32;
Query Match 74.1%; Score 43; DB 1; Length 177;
Best Local Similarity 62.5%; Pred. No. 1.32e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 31 DAPLDVRR 38
QY :|||:|
QY 222 EAPLNVS 229
RESULT 15
ID MKK2_CRILO STANDARD; PRT: 329 AA.
AC P49136;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE MAP KINASE-ACTIVATED PROTEIN KINASE 2 (EC 2.7.1.-) (MAPK-ACTIVATED
DE PROTEIN KINASE 2) (MAPKAP KINASE 2) (MAPKAP-2) (P45-54 HSP27 KINASE)
DE (FRAGMENT).
OS CRICETULUS LONGICAUDATUS (LONG-TAILED HAMSTER) (CHINESE HAMSTER).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; CRICETINAE; CRICETULUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RA HUOT J., LAMBERT H., LAVOIE J.N., GUIMOND A., HOULE F., LANDRY J.;
RL SUBMITTED (OCT-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: ITS PHYSIOLOGICAL SUBSTRATE SEEMS TO BE THE SMALL HEAT
CC SHOCK PROTEIN (HSP27/HSP25). IN VITRO CAN PHOSPHORYLATES GLYCOGEN
CC SYNTHASE AT SER-7 AND TYROSINE HYDROXYLASE (ON SER-19 AND SER-40).
CC THIS KINASE PHOSPHORYLATES SER IN THE PEPTIDE SEQUENCE, HYD-X-R-
CC X(2)-S, WHERE HYD IS A LARGE HYDROPHOBIC RESIDUE (BY SIMILARITY).
CC -!- ENZYME REGULATION: SEEMS TO BE ACTIVATED BY TWO DISTINCT PATHWAYS:
CC THE FIRST INVOLVES THE STIMULATION OF P42/P44 MAPK BY GROWTH
CC FACTORS, THE SECOND, TRIGGERED BY STRESS AND HEAT SHOCK, DEPENDS
CC ON THE ACTIVATION OF MPK2 AND UPSTREAM MAPKK/MAPKKK.
CC -!- PTM: PHOSPHORYLATED AND ACTIVATED BY MAP KINASE.
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DE EMBL: X82220; G559435; -
DR PROSITE: PS00107; PROTEIN_KINASE_ATP: PARTIAL.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM: PF00069; pkinase; 1.
DR HSSP: Q63450; 1A06.
KW TRANSFERASE: SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
RN PHOSPHORYLATION.
FT NON_TER 1 1
FT DOMAIN <1 254
FT NP_BIND <1 7
FT BINDING 22 22
ATP (BY SIMILARITY).
ATP (BY SIMILARITY).

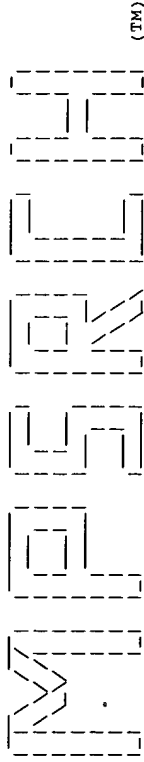
FT ACT_SITE 115 115 BY SIMILARITY.
FT MOD_RES 263 263 PHOSPHORYLATION (BY MAPK) (BY
FT SIMILARITY)
SQ SEQUENCE 329 AA; 38014 MW; 460A4727 CRC32;

Query Match 74.1%; Score 43; DB 1; Length 329;
Best Local Similarity 55.6%; Pred. NO. 1.32e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 262 QTPLHRSRV 270
::||: |||
Qy 222 EAPLNVSrv 230

Search completed: Thu Jul 8 18:33:21 1999
Job time : 7 secs.

THIS PAGE BLANK (USPTO)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:33:41 1999; MasPar time 7.89 Seconds
Tabular output not generated. 62.273 Million cell updates/sec

Title: >US-09-041-236-2
Description: (222-230) from US09041236.pep (14 of 45)
Perfect Score: 58
Sequence: 1 EAPLNSRV 9

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 21.726; Variance 20.401; scale 1.065

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	58	100.0	393	11	SEMAPHORIN L (FRAGMENT)	1.05e-03
2	58	100.0	666	4	SEMAPHORIN L	1.05e-03
3	48	82.8	545	2	PUTATIVE ACCESSORY PRO	7.51e-01
4	48	82.8	545	2	PUTATIVE ACCESSORY PRO	7.51e-01
5	47	81.0	413	11	HP-55.	1.39e+00
6	47	81.0	932	13	ACID ALPHA GLUCOSIDASE	1.39e+00
7	47	81.0	3099	14	POLYPROTEIN.	1.39e+00
8	46	79.3	326	11	CARBOXYL TERMINAL LIM	2.54e+00
9	46	79.3	329	4	LIM DOMAIN PROTEIN CLP	2.54e+00
10	45	77.6	102	6	SUCRASE-ISOMALTAZE, IN	4.60e+00
11	45	77.6	156	14	SMALL 1.	4.60e+00
12	45	77.6	233	1	ABC TRANSPORTER, ATP-B	4.60e+00
13	45	77.6	238	2	PUTATIVE DTPD-4-KETO-L	4.60e+00
14	45	77.6	378	2	PHASOLOTOXIN OPERON:	4.60e+00
15	45	77.6	427	6	GOR ANTIGENIC EPTOPE.	4.60e+00
16	45	77.6	658	2	HEAT SHOCK PROTEIN.	4.60e+00
17	44	75.9	236	11	MYELOID DIFFERENTIATIO	8.26e+00
18	44	75.9	348	2	FIMBRILIN.	8.26e+00
19	44	75.9	348	2	FIMBRILIN.	8.26e+00
20	44	75.9	350	2	FIMBRILIN.	8.26e+00

21	44	75.9	390	2	Q50822	DNA SEQUENCE, COMPLETE	8.26e+00
22	44	75.9	441	11	O64644	CHOLINE DEHYDROGENASE	8.26e+00
23	44	75.9	681	2	O53475	HYPOTHETICAL 74.9 KD P	8.26e+00
24	44	75.9	690	1	O57901	590AA LONG HYPOTHETICA	8.26e+00
25	44	75.9	818	14	O56051	STRUCTURAL POLYPROTEIN	8.26e+00
26	43	74.1	110	2	O56435	RIBOSOMAL PROTEIN L24.	1.47e+01
27	43	74.1	111	2	O84523	L24 RIBOSOMAL PROTEIN.	1.47e+01
28	43	74.1	305	1	O26838	CONSERVED PROTEIN.	1.47e+01
29	43	74.1	360	2	O51827	FIMBRILIN.	1.47e+01
30	43	74.1	413	11	O63969	SERINE PROTEASE INHIBI	1.47e+01
31	43	74.1	442	2	O34974	YTNJ.	1.47e+01
32	43	74.1	508	5	P90790	D2030.3 PROTEIN.	1.47e+01
33	43	74.1	760	5	O22235	T05E11.3 PROTEIN.	1.47e+01
34	43	74.1	768	10	O25638	HEAT SHOCK PROTEIN PRE	1.47e+01
35	43	74.1	769	13	O93497	ANDROGEN RECEPTOR.	1.47e+01
36	43	74.1	781	10	O43638	HEAT-SHOCK PROTEIN PRE	1.47e+01
37	43	74.1	808	6	O29091	PROTEIN KINASE.	1.47e+01
38	43	74.1	811	14	O84677	SIMILAR TO KLUVVEROMYC	1.47e+01
39	43	74.1	1142	4	O14324	FAST MYBP-C.	1.47e+01
40	43	74.1	2470	13	O90681	CATION-INDEPENDENT MAN	1.47e+01
41	42	72.4	332	14	O41951	HYPOTHETICAL 37.1 KD P	2.58e+01
42	42	72.4	413	11	O54759	ALPHA,-ANTITRYPSIN-LIK	2.58e+01
43	42	72.4	501	2	O82884	TYPE II SECRETION PATH	2.58e+01
44	42	72.4	1076	1	O28994	CARBAMOYL-PHOSPHATE SY	2.58e+01
45	42	72.4	1749	2	P73871	HYPOTHETICAL 190.2 KD	2.58e+01

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	393 AA.
ID O88371			
AC O88371;			
DT 01-NOV-1998 (TREMREL. 08, CREATED)			
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)			
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)			
DE SEMAPHORIN L (FRAGMENT).			
GN SEMAL.			
OS MUS MUSCULUS (MOUSE).			
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC EUCARYOTAC; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC EUCARYOTAC; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE; 98389619.			
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;			
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA			
RL viruses."			
RL GENOMICS 51:340-350(1998).			
DR EMBL; AF030699; G3523117; -			
FT NON_TER 393			
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;			
Query Match 100.0%; Score 58; DB 11; Length 393;			
Best Local Similarity 100.0%; Pred. No. 1.05e-03;			
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Db 252 EAPLNSRV 250			
Qy 222 EAPLNSRV 230			
RESULT 2	PRELIMINARY;	PRT;	666 AA.
ID O75326			
AC O75326;			
DT 01-NOV-1998 (TREMREL. 08, CREATED)			
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)			
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)			
DE SEMAPHORIN L.			
GN SEMAL.			
OS HOMO SAPIENS (HUMAN).			
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC CATARRHINI; HOMINIDAE; HOMO.			
RN [1]			
RP SEQUENCE FROM N.A.			

RX MEDLINE; 98399619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030698; G3523115; -.
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 58; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 254 EAPLNVSRV 262
| | | | | | | | | |
Qy 222 EAPLNVSRV 230

RESULT 3
ID Q48029 PRELIMINARY; PRT; 545 AA.
AC Q48029;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE ACCESSORY PROCESSING PROTEIN.
GN HMW2B.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-12;
RA MEDLINE; 92192797.
RX BARENKAMP S.J., LEININGER E.;

RT "Cloning, expression, and DNA sequence analysis of genes encoding
RT nontypeable Haemophilus influenzae high-molecular-weight
RT surface-exposed proteins related to filamentous hemagglutinin of
RT Bordetella pertussis."
RL INFECT. IMMUN. 60:1302-1313(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-12;
RX MEDLINE; 94314451.
RA BARENKAMP S.J., GEME J.W. III;
RT "Genes encoding high-molecular-weight adhesion proteins of
RT nontypeable Haemophilus influenzae are part of gene clusters."
RL INFECT. IMMUN. 62:3320-3328(1994).
DR EMBL; U08875; G482842; -.
SQ SEQUENCE 545 AA; 60786 MW; 71906899 CRC32;

Query Match 82.8%; Score 48; DB 2; Length 545;
Best Local Similarity 66.7%; Pred. No. 7.51e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 158 ENPLKVTRV 166
| | | | | | | | | |
Qy 222 EAPLNVSRV 230

RESULT 4
ID Q48032 PRELIMINARY; PRT; 545 AA.
AC Q48032;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE ACCESSORY PROCESSING PROTEIN.
GN HMW1B.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-12;
RX MEDLINE; 92192797.

RA BARENKAMP S.J., LEININGER E.;
RT "Cloning, expression, and DNA sequence analysis of genes encoding
RT nontypeable Haemophilus influenzae high-molecular-weight
RT surface-exposed proteins related to filamentous hemagglutinin of
RT Bordetella pertussis."
RL INFECT. IMMUN. 60:1302-1313(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-12;
RX MEDLINE; 94314451.
RA BARENKAMP S.J., GEME J.W. III;
RT "Genes encoding high-molecular-weight adhesion proteins of
RT nontypeable Haemophilus influenzae are part of gene clusters."
RL INFECT. IMMUN. 62:3320-3328(1994).
DR EMBL; U08876; G475772; -.
SQ SEQUENCE 545 AA; 60674 MW; 51168CCA CRC32;

Query Match 82.8%; Score 48; DB 2; Length 545;
Best Local Similarity 66.7%; Pred. No. 7.51e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 158 ENPLKVTRV 166
| | | | | | | | | |
Qy 222 EAPLNVSRV 230

RESULT 5
ID O54756 PRELIMINARY; PRT; 413 AA.
AC O54756;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HP-55.
OS TAMIAS SIBRICUS (SIBERIAN CHIPMUNK).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHA; SCIURIDAE; TAMIAS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 98094263.
RA TAKAMATSU N., KOJIMA M., TANIYAMA M., OHBA K., UEMATSU T., SEGAWA C.,
RA TSUTOU S., WATANABE M., KONDO J., KONDO N., SHIBA T.;
RT "Expression of multiple alapha-antitrypsin-like genes in hibernating
RT species of the squirrel family."
RL GENE 204:127-132(1997).
DR EMBL; AB000545; D1025329; -.
SQ SEQUENCE 413 AA; 45991 MW; 2BE3001A CRC32;

Query Match 81.0%; Score 47; DB 11; Length 413;
Best Local Similarity 66.7%; Pred. No. 1.39e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 343 DAPLTVSKV 351
| | | | | | | | | |
Qy 222 EAPLNVSRV 230

RESULT 6
ID O73626 PRELIMINARY; PRT; 932 AA.
AC O73626;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ACID ALPHA GLUCOSIDASE.
GN GAAI.
OS COTURNIX COTURNIX JAPONICA (JAPANESE QUAIL).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; COTURNIX.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 98201735.
RA KUNITA R., NAKABAYASHI O., WU J.Y., HAGIWARA Y., MIZUTANI M.,

RA PENNYBACKER M., CHEN Y.T., KIKUCHI T.;
RT "Molecular cloning of acid alpha-glucosidase cDNA of Japanese quail
RT (Coturnix coturnix japonica) and the lack of its mRNA in acid maltase
RL deficient quails";
RL BIOCHIM. BIOPHYS. ACTA 1362:269-278(1997).
DR ENBL; AB00967; D1026824;
DR PROSITE; PS00129; GLYCOSYL HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL HYDROL_F31_2; 1.
SQ SEQUENCE 932 AA; 104689 MW; 844B9287 CRC32;

Query Match 81.0%; Score 47; DB 13; Length 932;
Best Local Similarity 66.7%; Pred. No. 1.39e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 EPLEVSRV 198
QY 222 EAPLNSRV 230

RESULT 7
ID OS6075 PRELIMINARY; PRT; 3099 AA.
AC OS6075;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS PEANUT MOTTLE VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M;
RA FIASINSKI S., GONZALES R.A., CASSIDY B.G.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL; AF023848; G2739030; -
KW POLYPROTEIN.
FT CHAIN 1 322 PUTATIVE P1 PROTEIN
FT CHAIN 323 779 PUTATIVE HELPER-COMPONENT-PROTEINASE.
FT CHAIN 780 1128 PUTATIVE P3 PROTEIN.
FT CHAIN 1129 1180 PUTATIVE 6 KDA PROTEIN 1.
FT CHAIN 1181 1814 PUTATIVE CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN 1815 1867 PUTATIVE 6 KDA PROTEIN 2.
FT CHAIN 1868 2303 PUTATIVE NUCLEAR INCLUSION PROTEIN A.
FT CHAIN 2304 2821 PUTATIVE NUCLEAR INCLUSION PROTEIN B.
FT CHAIN 2822 3099 PUTATIVE COAT PROTEIN.
SQ SEQUENCE 3099 AA; 351034 MW; 8252CA4F CRC32;

Query Match 81.0%; Score 47; DB 14; Length 3099;
Best Local Similarity 75.0%; Pred. No. 1.39e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 110 APLHVARV 117
QY 223 APLNSRV 230

RESULT 8
ID O70400 PRELIMINARY; PRT; 326 AA.
AC O70400;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE CARBOXYL TERMINAL LIM DOMAIN PROTEIN.
GN LDB3.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MORIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA KOTAKA M., NGAI S.M., TSUI S.K.W., FUNG K.P., LEE C.Y., WAYE M.M.Y.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL; AF053367; G2996196; -
SQ SEQUENCE 326 AA; 35717 MW; 2883E487 CRC32;

Query Match 79.3%; Score 46; DB 11; Length 326;
Best Local Similarity 66.7%; Pred. No. 2.54e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 24 EQLPLASRV 32
QY 222 EAPLNSRV 230

RESULT 9
ID O00151 PRELIMINARY; PRT; 329 AA.
AC O00151;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE LIM DOMAIN PROTEIN CLP-36.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA KOTAKA M., NGAI S.M., GARCIA-BARCELO M., TSUI S.K.W., FUNG K.P.,
RA LEE C.Y., WAYE M.M.Y.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL; U90878; G1905874; -
DR PFAM; PF00412; LIM; 1.
DR PFAM; PF00595; PDZ; 1.
SQ SEQUENCE 329 AA; 36171 MW; 70864229 CRC32;

Query Match 79.3%; Score 46; DB 4; Length 329;
Best Local Similarity 66.7%; Pred. No. 2.54e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 25 EQLPLASRV 33
QY 222 EAPLNSRV 230

RESULT 10
ID Q95287 PRELIMINARY; PRT; 102 AA.
AC Q95287;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE SUCRASE-ISOMALTASE, INTESTINAL (FRAGMENT).
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA WINTEROE A.K., FREDHOLM M., DAVIES W.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL; Z81242; E276410; -
FT NON_TER 1 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11348 MW; 43D72BF5 CRC32;

Query Match 77.6%; Score 45; DB 6; Length 102;
Best Local Similarity 85.7%; Pred. No. 4.60e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 92 PLNVPRV 98
QY 224 PLNVSRV 230

RESULT 11
ID Q89790 PRELIMINARY; PRT; 156 AA.
AC Q89790;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

```
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SMALL 1.
OS THEILER'S ENCEPHALOMYELITIS VIRUS.
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC CARDIOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WM;
RX MEDLINE: 96130195.
RA MICHELIS T., JAROUSSE N., BRAHIC M.;
RT "Analysis of the leader and capsid coding regions of persistent and
RT neurovirulent strains of Theiler's virus.";
RL VIROLOGY 214:550-558(1995).
DR EMBL: U33046; G1237421; -.
DR EMBL: U33045; G1237418; -.
FT CHAIN 72 142 VP4.
FT CHAIN 143 156 VP2.
SQ SEQUENCE 156 AA; 17877 MW; A519D79C CRC32;

Query Match 77.6%; Score 45; DB 14; Length 156;
Best Local Similarity 55.68; Pred. No. 4.60e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 73 ETPHLTRV 81
|:|:|:|
QY 222 EAPLNVSRV 230

RESULT 12
ID Q28297 PRELIMINARY; PRT; 253 AA.
AC Q28297;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE ABC TRANSPORTER, ATP-BINDING PROTEIN.
GN AF1982.
OS ARCHAEOGLOBUS FULGIDUS.
OC ARCHAEOA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE;
OC ARCHAEOGLOBUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic,
RT sulphate-reducing archaeon Archaeoglobus fulgidus.";
RL NATURE 390:364-367(1997).
DR EMBL: AE000966; G2648557; -.
DR TIGR: AF1982; -.
DR PFAM: PF00005; ABC_tran; 1.
KW HYPOTHETICAL PROTEIN; ATP-BINDING.
SQ SEQUENCE 253 AA; 27616 MW; 60B0FADF CRC32;

Query Match 77.68; Score 45; DB 1; Length 253;
Best Local Similarity 77.88; Pred. No. 4.60e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 90 EAPLTVSOV 98
|:|:|:|
QY 222 EAPLNVSRV 230
```

```
RESULT 13
ID Q69119 PRELIMINARY; PRT; 298 AA.
AC Q69119;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE PUTATIVE DTDP-4-KETO-L-RHAMNOSE REDUCTASE.
GN RMLD.
OS BURKHOLDERIA PSEUDOMALLEI.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; BURKHOLDERIA GROUP;
OC BURKHOLDERIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1026B;
RA DESHAZER D., BRETT P.J., WOODS D.E.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF084070; G3135677; -.
SQ SEQUENCE 298 AA; 31613 MW; FDF62599 CRC32;

Query Match 77.6%; Score 45; DB 2; Length 298;
Best Local Similarity 75.0%; Pred. No. 4.60e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 241 APLAVSRI 248
|:|:|:|
QY 223 APLNVSRV 230

RESULT 14
ID Q52579 PRELIMINARY; PRT; 378 AA.
AC Q52579;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMREL. 01, LAST ANNOTATION UPDATE)
DE PHASELOTORIN OPERON: ORF1, ORF2, ORF3, ORF4, ORF5, AND ORF6 GENES,
DE COMPLETE CDS.
OS PSEUDOMONAS SYRINGAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC PSEUDOMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. PHASEOLICOLA;
RA ZHANG Y.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. PHASEOLICOLA;
RA SURESH P.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U27310; G1403595; -.
SQ SEQUENCE 378 AA; 41596 MW; A61881B2 CRC32;

Query Match 77.6%; Score 45; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 4.60e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 328 APLNVSR 334
|:|:|:|
QY 223 APLNVSR 229

RESULT 15
ID Q28509 PRELIMINARY; PRT; 427 AA.
AC Q28509;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE GOR ANTIGENIC EPITOPE.
OS PAN SP.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE.
RN [1]
RP SEQUENCE FROM N.A.
```

RX MEDLINE: 92126820.
 RA MISHIRO S., TAKEDA K., HOSHI Y., YOSHIKAWA A., GOTANDA T., ITOH Y.;
 RT "An autoantibody cross-reactive to hepatitis C virus core and a host
 nuclear antigen."
 RL AUTOIMMUNITY 10:269-273(1991).
 DR EMBL: S78897; G243898; -.
 DR PFAM: PF00929; Exonuclease: 1.
 SQ SEQUENCE 427 AA; 46564 MW; 966AA1B3 CRC32;

Query Match 77.6%; Score 45; DB 6; Length 427;
 Best Local Similarity 66.7%; Pred. No. 4.60e+00;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 397 EPPLPVPV 405
 Oy 222 EAPLNVSRV 230

Search completed: Thu Jul 8 18:33:54 1999
 Job time : 13 secs.

THIS PAGE BLANK (USPTO)

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:36:47 1999; MasPar time 4.34 Seconds
48.989 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2

Description: (228-237) from US09041236.pap (15 of 45)

Perfect Score: 74

Sequence: 1 SRVAQLCRGD 10

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 16.972; Variance 49.980; scale 0.340

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	71.6	861	32	W58540 Human semaphorin.	4.10e+01
2	53	71.6	861	22	W17658 Mouse CD100 antigen.	4.10e+01
3	52	70.3	974	33	W64221 Human secreted protei	5.29e+01
4	51	68.9	862	22	W17657 Human CD100 antigen.	6.81e+01
5	50	67.6	74	37	W74817 Human secreted protei	8.76e+01
6	50	67.6	317	18	W00165 Meiosis specific prot	8.76e+01
7	49	66.2	477	13	R74175 Human collapsin.	1.12e+02
8	49	66.2	712	13	R71394 Tribolium semaphorin	1.12e+02
9	49	66.2	771	13	R71380 Human semaphorin III	1.12e+02
10	49	66.2	887	25	W19856 Rat semaphorin Z.	1.12e+02
11	49	66.2	888	25	W19857 Human semaphorin Z.	1.12e+02
12	49	66.2	1618	11	R60127 Human nestin protein	1.12e+02
13	49	66.2	1618	5	R27205 Human nestin.	1.12e+02
14	48	64.9	28	7	R34560 Domain 1 from integra	1.44e+02
15	48	64.9	289	22	W15268 Salmonella secreted p	1.44e+02
16	48	64.9	775	36	W79193 Human Hrs-2 partial p	1.44e+02

17	48	64.9	924	36	W79192	1.44e+02
18	47	63.5	207	38	W88841 Polypeptide fragment	1.84e+02
19	47	63.5	219	37	W72890 Mycobacterium tubercu	1.84e+02
20	47	63.5	286	38	W88620 Secreted protein enco	1.84e+02
21	47	63.5	327	31	W48626 Human adenomatous pol	1.84e+02
22	47	63.5	411	38	W88846 Polypeptide fragment	1.84e+02
23	47	63.5	460	13	R69874 B thermoglucodasius D	1.84e+02
24	47	63.5	471	6	R29715 Heat stable hydantoin	1.84e+02
25	46	62.2	31	26	W21999 Insulin C chain pepti	2.36e+02
26	46	62.2	244	34	W77619 Mercuric reductase pr	2.36e+02
27	46	62.2	354	27	W30915 JC virus VP1 protein.	2.36e+02
28	46	62.2	354	26	W24016 Human JC virus VP1 pr	2.36e+02
29	46	62.2	423	32	W42111 Amino acid sequence o	2.36e+02
30	46	62.2	587	32	W51314 Human semaphorin W.	2.36e+02
31	46	62.2	776	32	W51313 Rat semaphorin W.	2.36e+02
32	45	60.8	29	13	R66179 Murine Flt3 ligand pe	3.00e+02
33	45	60.8	268	23	W01750 ECoRI binding fragmen	3.00e+02
34	45	60.8	278	9	R57414 Cytomegalovirus Towne	3.00e+02
35	45	60.8	373	18	R98461 Murine ICE-ced-3 homo	3.00e+02
36	45	60.8	373	31	W56031 Mouse ICH-3.	3.00e+02
37	45	60.8	373	13	R66767 Murine interleukin-1	3.00e+02
38	45	60.8	458	8	R41781 Rabbit HBI protein wh	3.00e+02
39	45	60.8	1014	35	W68474 HIV-1 strain YBF30 po	3.00e+02
40	44	59.5	201	38	W82405 Human NMLY-6 analogue	3.82e+02
41	44	59.5	460	39	W84181 A GDNFR-alpha-related	3.82e+02
42	44	59.5	502	38	W87561 CTLA4-p97 fusion prot	3.82e+02
43	44	59.5	502	38	W81585 CTLA4/p97 fusion prot	3.82e+02
44	44	59.5	708	38	W86022 Bovine lactoferrin (h	3.82e+02
45	44	59.5	708	31	W57318 Bovine lactoferrin.	3.82e+02

ALIGNMENTS

RESULT 1

ID W58540 standard; Protein; 861 AA.
AC W58540;
DT 02-SEP-1998 (first entry)
DE Human semaphorin.
KW Human; semaphorin; diagnosis; nervous disease; immune disease;
KW nerve extension inhibiting activity.
OS Homo sapiens.
PN J10155490-A.
PD 16-JUN-1998.
PF 27-NOV-1996; JP-332900.
PR (SUMU) SUMITOMO SEIYAKU KK.
PA WPI: 98-391044/34.
DR N-PSDB; V31121.
DR New human semaforin gene - useful in the diagnosis of nervous system
PT and immune disorders
PS Claim 1; Page 10-12; 15pp; Japanese.
CC The present sequence represents human semaphorin (translated from the
CC Japanese specification as semaforin) Semaphorin has nerve extension
CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
CC treatment and researches on nervous diseases and immune diseases.
SQ Sequence 861 AA;

Query Match 71.6%; Score 53; DB 32; Length 861;
Best Local Similarity 60.0%; Pred.No. 4.10e+01;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 251 prvarvckgd 260

:||||:|

QY 228 SRVAQLCRGD 237

RESULT 2

ID W17658 standard; Protein; 861 AA.

AC W17658;

DT 24-JUL-1997 (first entry)

DE Mouse CD100 antigen.

KW Cd100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;

KW vaccine.

```

OS Mus sp.
FH Key Location/Qualifiers
FT peptide 1..41
FT protein /label= Sig_peptide
FT protein 42..861
FT domain /label= Mat_protein
FT domain 42..553
FT domain /label= Semaphorin_domain
FT domain 554..630
FT domain /label= Ig-like_domain
FT domain 631..732
FT domain /label= Stalk_domain
FT domain 734..752
FT domain /label= Transmembrane_domain
FT domain 753..861
FT domain /label= Cytoplasmic_domain
FT modified_site 807..814
FT /label= Phosphorylation
FT /note= "putative tyrosine phosphorylation site"
PN WO9717368-A1.
PD 15-MAY-1997.
PF 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND ) DANA FARMER CANCER INST.
PI Bousiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB; T60666.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Example 8; Page 86-89; 135pp; English.
CC Mouse CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60666) isolated from murine T cells. Human CD100 antigen
CC (W17657) has also been identified. CD100 polypeptides and fusion
CC proteins, nucleic acids, and host cells expressing CD100 can be
CC utilised in diagnostic and therapeutic methods involving modulation
CC of B and T cell responses, neuron axonal growth and immune cell-
CC nerve cell interaction.
SQ Sequence 861 AA;

Query Match 71.6%; Score 53; DB 22; Length 861;
Best Local Similarity 60.0%; Pred. No. 4.10e+01;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 251 prvarvckgd 260
QY 228 SRVAQLCRGD 237
:||||:|

RESULT 3
ID W64221 standard; Protein: 974 AA.
AC W64221;
DE Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN WO9827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GENY ) GENETICS INST INC.
PI Agostino M, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-362774/31.

```

```

DR N-PSDB; V44295.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Claim 17j; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 70.3%; Score 52; DB 33; Length 974;
Best Local Similarity 60.0%; Pred. No. 5.29e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 251 prvarvcknd 260
QY 228 SRVAQLCRGD 237
:||||:|

RESULT 4
ID W17657 standard; Protein: 862 AA.
AC W17657;
DE 24-JUL-1997 (first entry)
KW Human CD100 antigen.
DE Human CD100 antigen.
KW vaccine.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..41
FT /label= Sig_peptide
FT protein 42..862
FT /label= Mat_protein
FT domain 42..553
FT /label= Semaphorin_domain
FT domain 554..630
FT /label= Ig-like_domain
FT domain 631..733
FT /label= Stalk_domain
FT domain 735..752
FT /label= Transmembrane_domain
FT domain 753..862
FT /label= Cytoplasmic_domain
FT modified_site 808..815
FT /label= Phosphorylation
FT /note= "putative tyrosine phosphorylation site"
PN WO9717368-A1.
PD 15-MAY-1997.
PF 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND ) DANA FARMER CANCER INST.
PI Bousiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB; T60665.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Claim 7; Page 70-72; 135pp; English.
CC Human CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60665) isolated by COS cell expression cloning. CD100
CC polypeptides and fusion proteins, nucleic acids, and host cells
CC expressing CD100 can be utilised in diagnostic and therapeutic
CC methods involving modulation of B and T cell responses, neuron
CC axonal growth and immune cell-nerve cell interaction.
SQ Sequence 862 AA;

```

Query Match 68.9%; Score 51; DB 22; Length 862;
Best Local Similarity 50.08; Pred.No. 6.81e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 251 priarvckgd 260
QY 228 SRVAQLCRGD 237
:!:!:!!

RESULT 5
ID W74817 standard; Protein; 74 AA.
AC W74817;
DT 25-JAN-1999 (first entry)
DE Human secreted protein encoded by gene 89 clone HNEBI60.
KW Human; secreted protein; testis; tumour; foetal brain tissue;
KW fusion protein; cancer; central nervous system; seizure;
KW diagnosis; neurodegenerative disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc-difference 74
FN /label= unknown

FN W09839448-A2.

PD 11-SEP-1998. U04493.

PF 06-MAR-1998; US-061060.

PR 02-OCT-1997; US-038621.

PR 07-MAR-1997; US-040161.

PR 07-MAR-1997; US-040162.

PR 07-MAR-1997; US-040163.

PR 07-MAR-1997; US-040333.

PR 07-MAR-1997; US-040334.

PR 07-MAR-1997; US-040336.

PR 07-MAR-1997; US-040626.

PR 11-APR-1997; US-043311.

PR 11-APR-1997; US-043312.

PR 11-APR-1997; US-043313.

PR 11-APR-1997; US-043314.

PR 11-APR-1997; US-043568.

PR 11-APR-1997; US-043569.

PR 11-APR-1997; US-043576.

PR 11-APR-1997; US-043578.

PR 11-APR-1997; US-043580.

PR 11-APR-1997; US-043669.

PR 11-APR-1997; US-043670.

PR 11-APR-1997; US-043671.

PR 11-APR-1997; US-043672.

PR 11-APR-1997; US-043674.

PR 23-MAY-1997; US-047492.

PR 23-MAY-1997; US-047500.

PR 23-MAY-1997; US-047501.

PR 23-MAY-1997; US-047502.

PR 23-MAY-1997; US-047503.

PR 23-MAY-1997; US-047581.

PR 23-MAY-1997; US-047582.

PR 23-MAY-1997; US-047583.

PR 23-MAY-1997; US-047584.

PR 23-MAY-1997; US-047585.

PR 23-MAY-1997; US-047586.

PR 23-MAY-1997; US-047587.

PR 23-MAY-1997; US-047588.

PR 23-MAY-1997; US-047589.

PR 23-MAY-1997; US-047590.

PR 23-MAY-1997; US-047592.

PR 23-MAY-1997; US-047593.

PR 23-MAY-1997; US-047594.

PR 23-MAY-1997; US-047595.

PR 23-MAY-1997; US-047596.

PR 23-MAY-1997; US-047597.

PR 23-MAY-1997; US-047598.

PR 23-MAY-1997; US-047599.

PR 23-MAY-1997; US-047600.

PR 23-MAY-1997; US-047601.

PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 13-JUN-1997; US-049610.
PR 08-JUL-1997; US-051926.
PR 16-JUL-1997; US-052874.
PR 18-AUG-1997; US-055724.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057659.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PR (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kyaw H, Larleau DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
PI WPI: 98-506364/43.
DR N-PSDB: V59599.
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 595; 721pp; English.
CC This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 89 from the human cDNA clone HNEBI60
CC (deposited as clone ATCC 97901 and ATCC 209047).
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. V59502) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample

CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see V59511 for described
CC uses).
SQ Sequence 74 AA;

Query Match 67.6%; Score 50; DB 37; Length 74;
Best Local Similarity 75.0%; Pred. No. 8.76e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 24 hvaplcrq 31
:|:|:|:|
QY 229 RVAQLCRG 236

RESULT 6

ID W00165 standard; Protein; 317 AA.
AC W00165;
DT 16-OCT-1996 (first entry)
DE Meiosis specific protein #1 encoded by AR401.
KW Transformation; host plant; meiosis; spore formation; yeast.
OS Synthetic.
PN J08140586-A.
PD 04-JUN-1996.
PF 25-NOV-1994; 315514.
PR 25-NOV-1994; JP-315514.
PA (TOVT) TOYOTA JIDOSHA KK.
DR WPI: 96-316320/32.
DR N-PSDB: T37084.

PT Gene coding for plant protein - useful in meiosis and spore
PT formation

PS Claim 6; Page 17-19; 26pp; Japanese.
CC The sequences given in W00160-66 are encoded by cDNA sequences which
CC were used to transform host plants. These protein sequences are
CC involved in physical activities in plants such as meiosis, etc.
CC in plant cells. The genes are useful for induction of meiosis and
CC spore formation after introduction into a yeast incapable of meiosis
CC and spore formation.
SQ Sequence 317 AA;

Query Match 67.6%; Score 50; DB 18; Length 317;
Best Local Similarity 66.7%; Pred. No. 8.76e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 297 rksnclcrq 305
:|:|:|:|
QY 229 RVAQLCRG 237

RESULT 7

ID R74175 standard; Protein; 477 AA.

AC R74175;
DT 01-NOV-1995 (first entry)
DE Human collapsin.
KW Collapsin; antibody; therapy.
OS Homo sapiens.

FH Key Location/Qualifiers
FT binding_site 9..19 "antibody binding site"
FT binding_site 51..65
FT /note: "antibody binding site"

US5416197-A.

PN 16-MAY-1995.

PF 15-OCT-1993; 136922.

PR 15-OCT-1993; US-136922.

PA (UYPE-) UNIV PENNSYLVANIA.

PI Luo Y, Raper JA;

DR WPI: 95-193478/25.

DR N-PSDB: Q92331.

PT New antibody to human collapsin - used to inhibit the activity of
PT collapsin, to induce neurite out-growth and to treat individuals with
PT nerve damage.
PS Claim 2; Columns 15-18; 11pp; English.

CC An antibody capable of specifically binding at least a portion of
CC the collapsin protein can be used to purify human collapsin and
CC to inhibit the activity of the protein. It can be used to induce
CC neurite outgrowth by neuronal cells and to treat individuals
CC suffering from nerve damage.
SQ Sequence 477 AA;

Query Match 66.2%; Score 49; DB 13; Length 477;
Best Local Similarity 40.0%; Pred. No. 1.12e+02;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 31 arigqicknd 40
:|:|:|:|
QY 228 SRVAQLCRGD 237

RESULT 8

ID R71384 standard; Protein; 712 AA.

AC R71384;

DT 21-NOV-1995 (first entry)

DE Tribolium semaphorin I protein.

KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;

KW variola major virus; smallpox; semaphorin receptor binding activity;

KW modulation; nerve cell growth; immune response; viral pathogenesis;

KW neurological disease; neuro-regeneration; oncological infection.

OS Tribolium sp.

PN W09507706-A.

PD 23-MAR-1995.

PF 13-SEP-1994; U10151.

PR 13-SEP-1993; US-121713.

PA (REGC) UNIV CALIFORNIA.

PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;

PI O'Connor T;

DR WPI: 95-131177/17.

DR N-PSDB: Q87446.

PT New class of semaphorin peptide(s) and polypeptide(s) - are

PT potent modulators of nerve cell growth and regeneration

PS Example 2; Page 85-88; 101pp; English.

CC The sequence of the beetle Tribolium semaphorin I protein. The gene was
CC isolated by PCR using Tribolium genomic DNA. The proteins encoded by the
CC grasshopper semaphorin I (Q87441), human semaphorin III (Q87442),
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I or variola major (smallpox) virus
CC semaphorin IV (Q87447) genes were used to generate a series of peptides
CC (R70370-R70418), which retain semaphorin receptor binding activity. The
CC semaphorin derived or semaphorin receptor derived peptides are potent
CC modulators of nerve cell growth, immune responsiveness and viral
CC pathogenesis. They can be used in diagnosis and treatment of neurological
CC disease and neuro-regeneration, immune modulation and diagnosis and
CC treatment of viral and oncological infection and diseases.

SQ Sequence 712 AA;

Query Match 66.2%; Score 49; DB 13; Length 712;
Best Local Similarity 60.0%; Pred. No. 1.12e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 236 srvarvckdd 245
|:|:|:|:|
QY 228 SRVAQLCRGD 237

RESULT 9

ID R71380 standard; Protein; 771 AA.

AC R71380;

DT 21-NOV-1995 (first entry)

DE Human semaphorin III protein.

KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;

KW variola major virus; smallpox; semaphorin receptor binding activity;

KW modulation; nerve cell growth; immune response; viral pathogenesis;

KW neurological disease; neuro-regeneration; oncological infection.

OS Homo sapiens.

PN W09507706-A.

PD 23-MAR-1995.

PF 13-SEP-1994; U10151.
 PR 13-SEP-1993; US-121713.
 PA (REGC) UNIV CALIFORNIA.
 PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
 DR O'Connor T;
 DR WPI: 95-131177/17.
 DR N-PSDB: Q87442.
 PT New class of semaphorin peptide(s) and polypeptide(s) - are
 PT potent modulators of nerve cell growth and regeneration
 PS Example 2; Page 60-63; 106pp; English.
 CC The sequence of the human semaphorin III protein. The proteins
 CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
 CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
 CC (Q87444-5), Tribolium semaphorin I (Q87446) or varicella major (smallpox)
 CC virus semaphorin IV (Q87447) genes were used to generate a series of
 CC peptides (R70370-R70418), which retain semaphorin receptor binding
 CC activity. The semaphorin derived or semaphorin receptor derived peptides
 CC are potent modulators of nerve cell growth, immune responsiveness and
 CC viral pathogenesis. They can be used in diagnosis and treatment of
 CC neurological disease and neuro-regeneration, immune modulation and
 CC diagnosis and treatment of viral and oncological infection and diseases.
 SQ Sequence 771 AA;

Query Match 66.2%; Score 49; DB 13; Length 771;
 Best Local Similarity 40.0%; Pred. No. 1.12e+02;
 Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 263 arigqicknd 272
 :|:|:|:|
 Qy 228 SRVAQLCRGD 237

RESULT 10
 ID W19856 standard; Protein: 887 AA.
 AC W19856;

DE 13-FEB-1998 (first entry)
 DE Rat semaphorin Z.
 KW Semaphorin Z; central nerve extension; rat; human; inhibitor;
 KW central nerve regeneration promoter.
 OS Rattus norvegicus.
 PN W09720928-A1.
 PD 12-JUN-1997.
 PF 02-DEC-1996; J03517.
 PR 31-OCT-1996; JP-307205.
 PR 06-DEC-1995; JP-345187.
 PA (SUMU) SUMITOMO PHARM CO LTD.
 PI Kikuchi K, Kimura T;
 DR WPI: 97-319775/29.
 DR N-PSDB: T72107.
 PT Semaphorin Z and the gene encoding it - also inhibitors of its
 PT action which can be used as promoters of central nerve regeneration
 PS Claim 3; Page 66-71; 106pp; Japanese.
 CC This sequence represents the rat semaphorin Z. This protein sequence, and
 CC the human semaphorin Z protein (see W19857) can be used for screening
 CC possible candidates for activity as semaphorin Z inhibitors. Proteins
 CC containing all or part of the semaphorin Z sequence are useful as central
 CC nerve extension inhibitors. Semaphorin Z inhibitors identified by
 CC screening can be used as promoters of central nerve regeneration.
 SQ Sequence 887 AA;

Query Match 66.2%; Score 49; DB 25; Length 887;
 Best Local Similarity 60.0%; Pred. No. 1.12e+02;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 262 srvarvcknd 271
 ||||:|:|:|
 Qy 228 SRVAQLCRGD 237

RESULT 11
 ID W19857 standard; Protein: 888 AA.
 AC W19857;
 DT 13-FEB-1998 (first entry)

DE Human semaphorin Z.
 KW Semaphorin Z; central nerve extension; rat; human; inhibitor;
 KW central nerve regeneration promoter.
 OS Homo sapiens.
 PN W09720928-A1.
 PD 12-JUN-1997.
 PF 02-DEC-1996; J03517.
 PR 31-OCT-1996; JP-307205.
 PR 06-DEC-1995; JP-345187.
 PA (SUMU) SUMITOMO PHARM CO LTD.
 PI Kikuchi K, Kimura T;
 DR WPI: 97-319775/29.
 DR N-PSDB: T72108.
 PT Semaphorin Z and the gene encoding it - also inhibitors of its
 PT action which can be used as promoters of central nerve regeneration
 PS Claim 3; Page 77-82; 106pp; Japanese.
 CC This sequence represents the human semaphorin Z. This protein sequence,
 CC and the rat semaphorin Z protein (see W19856) can be used for screening
 CC possible candidates for activity as semaphorin Z inhibitors. Proteins
 CC containing all or part of the semaphorin Z sequence are useful as central
 CC nerve extension inhibitors. Semaphorin Z inhibitors identified by
 CC screening can be used as promoters of central nerve regeneration.
 SQ Sequence 888 AA;

Query Match 66.2%; Score 49; DB 25; Length 888;
 Best Local Similarity 60.0%; Pred. No. 1.12e+02;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 261 srvarvcknd 270
 ||||:|:|:|
 Qy 228 SRVAQLCRGD 237

RESULT 12
 ID R60127 standard; Protein: 1618 AA.
 AC R60127;
 DT 21-MAR-1995 (first entry)
 DE Human nestin protein is useful to identify brain tumours.
 KW nestin gene; brain tumour; neoplastic cells; glial; neuronal;
 KW muscle; neural multipotential stem cell; mammalian brain; detection;
 KW diagnosis; medulloblastoma; glioblastoma; oligodendroglioma.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 1235..1236
 FT /note= "encoded by CTCAG"
 FT misc_difference 1341..1342
 FT /note= "encoded by GACCC"
 PN US5338839-A.
 PD 16-AUG-1994.
 PF 12-APR-1988; 180548.
 PR 12-APR-1988; US-180548.
 PR 02-JUN-1988; US-201762.
 PR 25-OCT-1990; US-603803.
 PR 22-FEB-1991; US-660412.
 PR 19-MAR-1992; US-853913.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Lendahl U, McKay RDG;
 DR WPI: 94-363332/32.
 DR N-PSDB: Q70448.
 PT Nucleotide and protein sequences for human and rat nestin -
 PT distinguishes neural multipotential stem cells and brain tumour
 PT cells from more differentiated cell types; for use in the
 PT diagnosis of brain tumours
 PS Claim 5; Column 47-56; 45pp; English.
 CC This sequence is the human nestin protein encoded by Q70448.
 CC Nestin protein expression distinguishes neural multipotential stem
 CC cells and brain tumour cells from the more differentiated neural
 CC cell types (eg., neuronal, glial and muscle cells of the adult brain).
 CC The nestin protein can be used in diagnosing tumours of the brain,
 CC such as medulloblastomas, glioblastomas and oligodendroglioma.
 CC (See also R60126).
 SQ Sequence 1618 AA;

Query Match 66.2%; Score 49; DB 11; Length 1618;
Best Local Similarity 60.0%; Pred. No. 1.12e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 486 srfsicrge 495
||| :|||
QY 228 SRVAQLCRGD 237

RESULT 13
ID R27205 standard; Protein; 1618 AA.
AC R27205;
DT 15-FEB-1993 (first entry)
DE Human nestin.
KW Intermediate filament; central nervous system; brain tumour;
KW neurofilament.
OS Homo sapiens.
PN WO9214821-A.
PD 03-SEP-1992.
PF 21-FEB-1992; U01375.
PR 22-FEB-1991; US-660412.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Lendahl U, McKay RDG;
DR WPI: 92-316175/38.
DR N-PSDB: Q28399.
PT Diagnosis of pre-disposition to brain tumours - using DNA
PT encoding nestin which distinguishes neural multi-potential stem
PT cells from neuronal, glial and muscle cells
PS Claim 3: Fig 2: 63pp; English.
CC The human nestin gene was isolated using low stringency DNA
CC hybridisation of a human genomic phage lambda library with a rat
CC nestin probe. The amino acid sequence of the nestin gene was
CC deduced from the nucleotide sequence. The human nestin amino acid
CC sequence shows 75% similarity with the rat nestin sequence. There
CC is more than 60% identity between the two sequences. Antibodies to
CC nestin protein can be used in vivo diagnosis of brain tumours.
CC See also Q28398 for the rat nestin gene.
SQ Sequence 1618 AA;

Query Match 66.2%; Score 49; DB 5; Length 1618;
Best Local Similarity 60.0%; Pred. No. 1.12e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 486 srfsicrge 495
||| :|||
QY 228 SRVAQLCRGD 237

RESULT 14
ID R34560 standard; Protein; 28 AA.
AC R34560;
DT 04-JUN-1993 (first entry)
DE Domain 1 from integrase from phage HK22.
KW Cytotoxic T-lymphocyte response; transformed Mycobacteria; BCG;
KW Mycobacterium smegmatis; vaccine; cell mediated immunity; HIV;
KW Pertussis; malaria; Influenza virus; CTL; herpes virus.
PN Bacteriophage HK22.
PN WO9307897-A.
PD 29-APR-1993.
PF 21-OCT-1992; U09075.
PR 21-OCT-1991; US-780261.
PA (MEDI-) MEDIMUNE INC.
PI Stover CK;
DR WPI: 93-152187/18.
PT Expression vector for expressing protein or polypeptide in
PT mycobacterium - contg DNA sequences encoding lipoprotein
PT secretion signal and peptide heterologous to bacteria
PT fusion protein of lipoprotein heterologous to bacteria
PS Disclosure; Fig 29; 86pp; English.
CC This sequence is domain 1 from the integrase of phage HK22. The
CC whole coding sequence of the enzyme could be included in an expression
CC vector in order to aid its integration into its host. The
CC expression vector also includes at least the secretion signal of a

CC lipoprotein and a second sequence encoding a heterologous protein
CC and a mycobacterial promoter to control expression of the
CC heterologous protein.
SQ Sequence 28 AA;

Query Match 64.9%; Score 48; DB 7; Length 28;
Best Local Similarity 71.4%; Pred. No. 1.44e-02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

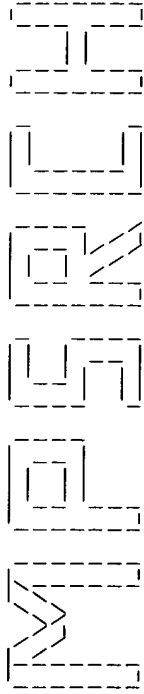
Db 13 rvgdlcr 19
||:||||
QY 229 RVAQLCR 235

RESULT 15
ID W15268 standard; Protein; 289 AA.
AC W15268;
DT 04-AUG-1997 (first entry)
DE Salmonella secreted protein SspB (incomplete).
KW Salmonella secreted protein; SspB; bacterial-mediated endocytosis;
KW diagnosis; therapy; vaccine; attenuation; virulence.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT Protein 1..289
FT /note= "amino acids 96-117 of the SspB amino acid
FT sequence of Fig 22 do not correspond to the
FT translated sequence of the SspB gene of
FT Fig 19"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GENO) GEN HOSPITAL CORP.
PI Miller SI.
DR WPI: 97-289217/26.
DR N-PSDB: T67028.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 3: Fig 22: 95pp; English.
CC Salmonella typhimurium secreted proteins SspB, SspC, SspD and SspA
CC (W15268-71) are virulence factors involved (except SspA) in
CC bacterial mediated endocytosis, and are able to translocate into
CC human epithelial cells. Ssp secretion is dependent on expression
CC of PrgH (see also T67036). Ssp genes (T67028-31) were isolated
CC by examining S. typhimurium mutants in invasion assays on HEp-2
CC epithelial cells. Ssp polypeptides can be used in methods for the
CC detection, treatment or prevention of Salmonella infection.
CC Chimeric proteins incorporating Ssp can be used to target an
CC antigen to epithelial cells or to induce a cytotoxic T cell
CC response in a mammal. Decreased secretion of Ssp provides
CC attenuated bacterial cells suitable as live vaccines.
SQ Sequence 289 AA;

Query Match 64.9%; Score 48; DB 22; Length 289;
Best Local Similarity 55.6%; Pred. No. 1.44e-02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 108 rlsplchgd 116
||:||||
QY 229 RVAQLCRGD 237

Search completed: Thu Jul 8 18:37:26 1999
Job time : 39 secs.



(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:36:15 1999; Maspar time 5.10 Seconds
Tabular output not generated. 78.605 Million cell updates/sec

Title: >US-09-041-236-2
Description: (228-237) from US09041236.pep (15 of 45)
Perfect Score: 74
Sequence: 1 SRVAQLCRGD 10

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.986; Variance 30.380; scale 0.757

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	58	78.4	489	2 F65125 probable general secr	2.22e-01
2	58	78.4	748	2 I48744 semaphorin A - mouse	2.22e-01
3	58	78.4	749	2 G01856 semaphorin V - human	2.22e-01
4	56	75.7	161	2 T00741 acyl carrier protein,	5.78e-01
5	56	75.7	605	1 W2BE33 gene 33 protein - hum	5.78e-01
6	56	75.7	782	2 I48746 semaphorin C - mouse	5.78e-01
7	55	74.3	354	2 S75877 hypothetical protein	9.26e-01
8	55	74.3	760	2 I48745 semaphorin B - mouse	9.26e-01
9	54	73.0	834	2 S66498 M-sema F protein prec	1.48e+00
10	53	71.6	1215	2 T00364 hypothetical protein	2.34e+00
11	53	71.6	1827	2 A35694 cut1 protein - fissio	2.34e+00
12	51	68.9	425	2 C70327 serine--trna ligase (5.80e+00
13	51	68.9	430	2 S75313 serine--trna ligase (5.80e+00
14	50	67.6	173	2 E71017 probable NADH-ubiquin	9.05e+00
15	50	67.6	423	2 S43135 serine--trna ligase (9.05e+00
16	50	67.6	1137	2 G70868 probable regulatory p	9.05e+00
17	50	67.6	1884	2 Jc4975 plexin 2 precursor -	9.05e+00
18	49	66.2	250	2 F70741 hypothetical protein	1.40e+01
19	49	66.2	666	2 I58169 semaphorin III - mous	1.40e+01
20	49	66.2	711	2 A49423 semaphorin I precursor	1.40e+01
21	49	66.2	771	2 D49423 semaphorin III precu	1.40e+01
22	49	66.2	772	2 I48747 semaphorin D - mouse	1.40e+01
23	49	66.2	772	2 A49069 collapsin - chicken	1.40e+01

24	49	66.2	1618	2 S21424 nestin - human	1.40e+01
25	48	64.9	313	2 D69945 phage-related protein	2.17e+01
26	48	64.9	357	2 S04990 integrase - phage HK0	2.17e+01
27	48	64.9	431	2 JWO098 carbazole dioxygenase	2.17e+01
28	48	64.9	446	2 S69051 hypothetical protein	2.17e+01
29	48	64.9	446	2 H65060 glucarate dehydratase	2.17e+01
30	48	64.9	543	1 KIHUPL pyruvate kinase (EC 2	2.17e+01
31	48	64.9	574	1 KIHUPR pyruvate kinase (EC 2	2.17e+01
32	48	64.9	581	2 I45889 ezrin - bovine	2.17e+01
33	48	64.9	586	1 A34400 ezrin - human	2.17e+01
34	48	64.9	586	1 B41129 ezrin - mouse	2.17e+01
35	48	64.9	1055	2 Jc5216 type I site-specific	3.32e+01
36	47	63.5	219	2 A70734 probable cutinase pre	3.32e+01
37	47	63.5	266	2 S60674 hypothetical protein	3.32e+01
38	47	63.5	327	2 G01037 t-cell activation pro	3.32e+01
39	47	63.5	471	2 Jc2310 dihydropyrimidinase (3.32e+01
40	47	63.5	506	2 S67256 probable pyruvate kin	3.32e+01
41	47	63.5	552	2 I54388 LZTR-1 - human	3.32e+01
42	47	63.5	706	1 W2BED8 gene 45 protein - equ	3.32e+01
43	47	63.5	875	2 F70755 hypothetical protein	3.32e+01
44	47	63.5	1010	2 A33509 vinculin - Caenorhabd	3.32e+01
45	47	63.5	1085	2 E70834 probable regulatorypr	3.32e+01

ALIGNMENTS

RESULT 1
ENTRY F65125 #type complete
TITLE probable general secretion pathway protein b - Escherichia coli (strain K-12)
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 14-Nov-1997
ACCESSIONS F65125
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Colliado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession F65125
#status preliminary; nucleic acid sequence not shown;

##molecule_type DNA
##translation not shown

##residues 1-489 ##label BLAT
##cross-references GB:AE000409; GB:U00096; NID:gl789718; PID:gl789720;
#experimental_source strain K-12, substrain MG1655

GENETICS
#gene yheD

SUMMARY
#length 489 #molecular-weight 55297 #checksum 9220

Query Match 78.4%; Score 58; DB 2; Length 489;

Best Local Similarity 77.8%; Pred. No. 2.22e-01;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 186 RIALLCRGD 194

QY 229 RVAQLCRGD 237

RESULT 2

ENTRY I48744 #type complete

TITLE semaphorin A - mouse

ORGANISM #formal_name Mus musculus #common_name house mouse

DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 04-Sep-1998

ACCESSIONS I48744

REFERENCE I48744

#authors Puschel, A.W.; Adams, R.H.; Betz, H.

```
#journal      Neuron (1995) 14:941-948
#title        Murine semaphorin D/collapsin is a member of a diverse gene
               family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession    I48744
#status       preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues    1-748 ##label RES
##cross-references EMBL:X85990; NID:g854323; PID:g854324
GENETICS
#gene         semA
CLASSIFICATION #superfamily semaphorin
SUMMARY       #length 748 #molecular-weight 82894 #checksum 9017
               78.4%; Score 58; DB 2; Length 748;
Query Match   70.0%; Pred. No. 2.22e-01;
Best Local Similarity
Matches       7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 262 SRVQICRND 271
   |||:||||
QY 228 SRVAQLCRGD 237

RESULT 3
ENTRY   G01856           #type complete
TITLE   semaphorin V - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE    21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
04-Sep-1998
ACCESSIONS G01856
REFERENCE   G08634
#authors    Sekido, Y.
#submission submitted to the EMBL Data Library, June 1995
#accession  G01856
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues  1-749 ##label SEK
##cross-references EMBL:U28369; NID:g974283; PID:g974284
CLASSIFICATION #superfamily semaphorin
SUMMARY       #length 749 #molecular-weight 83121 #checksum 2747

Query Match   78.4%; Score 58; DB 2; Length 749;
Best Local Similarity 70.0%; Pred. No. 2.22e-01;
Matches       7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 263 SRVQICRND 272
   |||:||||
QY 228 SRVAQLCRGD 237

RESULT 4
ENTRY   T00741           #type fragment
TITLE   acyl carrier protein, mitochondrial - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE    26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change
26-Feb-1999
ACCESSIONS T00741
REFERENCE   Z14201
#authors    Adams, M.D.; Loftus, B.J.; Zhou, L.; Crosby, M.; Fuhrmann,
               J.; Brandon, R.; Kim, U.J.; Kerlavage, A.R.; Venter, J.C.
#submission submitted to the EMBL Data Library, March 1998
#description Human Chromosome 16 BAC clone CIT987SK-A-735G6.
#accession  T00741
#status     translated from GB/EMBL/DBJ
#molecule_type DNA
##residues  1-161 ##label ADA
##cross-references EMBL:AC002400; NID:g2576344; PID:g2576345
GENETICS
#map_position 16p12
#genome      nuclear
#note        Intron positions not resolved (incomplete sequence)
A-735G6.1
KEYWORDS     mitochondrion
```

```
SUMMARY       #length 161 #checksum 7276

Query Match   75.7%; Score 56; DB 2; Length 161;
Best Local Similarity 75.0%; Pred. No. 5.78e-01;
Matches       6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 64 GRVTQLCR 71
   |||:||||
QY 228 SRVAQLCR 235

RESULT 5
ENTRY   W2BE33           #type complete
TITLE   gene 33 protein - human herpesvirus 3
ORGANISM #formal_name human herpesvirus 3, varicella-zoster virus
DATE    30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change
14-Nov-1997
ACCESSIONS A27214
REFERENCE   A27345
#authors    Davison, A.J.; Scott, J.E.
#journal    J. Gen. Virol. (1986) 67:1759-1816
#title      The complete DNA sequence of varicella-zoster virus.
#cross-references MUID:86306657
#accession  G27214
#molecule_type DNA
##residues  1-605 ##label DAV
##cross-references EMBL:X04370; NID:g59989; PID:g60022
GENETICS
#gene       33
CLASSIFICATION #superfamily varicella-zoster virus gene 33 protein
SUMMARY       #length 605 #molecular-weight 66046 #checksum 8450

Query Match   75.7%; Score 56; DB 1; Length 605;
Best Local Similarity 66.7%; Pred. No. 5.78e-01;
Matches       6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 293 KVTQLCRND 301
   ||:|||||
QY 229 RVAQLCRGD 237

RESULT 6
ENTRY   I48746           #type fragment
TITLE   semaphorin C - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE    02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
28-Feb-1997
ACCESSIONS I48746
REFERENCE   I48744
#authors    Puschel, A.W.; Adams, R.H.; Betz, H.
#journal    Neuron (1995) 14:941-948
#title      Murine semaphorin D/collapsin is a member of a diverse gene
               family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession  I48746
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues  1-782 ##label RES
##cross-references EMBL:X85992; NID:g854327; PID:g854328
GENETICS
#gene       semC
SUMMARY       #length 782 #checksum 1571

Query Match   75.7%; Score 56; DB 2; Length 782;
Best Local Similarity 70.0%; Pred. No. 5.78e-01;
Matches       7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 226 SRVARVCKGD 235
   ||||:||||
QY 228 SRVAQLCRGD 237

RESULT 7
```



```

ENTRY      S75877      #type complete
TITLE      hypothetical protein sir1166 - Synecchocystis sp. (strain PCC
ORGANISM    #6803)
#variety    #formal_name Synecchocystis sp.
DATE        25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
ACCESSIONS PCC 6803
REFERENCE   S75877
#authors    Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
            Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugiyura, M.;
            Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
            Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
            S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
            Yasuda, M.; Tabata, S.
#journal    DNA Res. (1996) 3:109-136
#title      Sequence analysis of the genome of the unicellular
            cyanobacterium Synecchocystis sp. PCC6803. II. Sequence
            determination of the entire genome and assignment of
            potential protein-coding regions.
#cross-references MUID:97061201
#accession  S75877
#status     nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues   1-354 #label KAN
#cross-references EMBL:D80913; GB:AB001339; NID:g1653348; PID:d1019069;
            PID:g1653422
#note       the nucleotide sequence was submitted to the EMBL Data
            Library, June 1996
SUMMARY     #length 354 #molecular-weight 39290 #checksum 2108

Query Match 74.3%; Score 55; DB 2; Length 354;
Best Local Similarity 77.8%; Pred. No. 9.26e-01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 137 NRVAQLCPG 145
QY 228 SRVAQLCRG 236
|||||

RESULT      8
ENTRY      I48745      #type complete
TITLE      semaphorin B, mouse
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
ACCESSIONS I48745
REFERENCE   I48745
#authors    Paschel, A.W.; Adams, R.H.; Betz, H.
#journal    Neuron (1995) 14:941-948
#title      Murine semaphorin D/collapsin is a member of a diverse gene
            family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession  I48745
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues   1-760 #label RES
#cross-references EMBL:X85991; NID:g854325; PID:g854326
GENETICS    semB
#gene       #length 760 #molecular-weight 83458 #checksum 2188
SUMMARY

Query Match 74.3%; Score 55; DB 2; Length 760;
Best Local Similarity 70.0%; Pred. No. 9.26e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 263 SRVAQCKND 272
QY 228 SRVAQLCRG 237
|||||

RESULT      9
ENTRY      S66498      #type complete

```

```

TITLE      M-sema F protein precursor - mouse
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change
ACCESSIONS S66498
REFERENCE   S66498
#authors    Inagaki, S.; Furuyama, T.; Iwahashi, Y.
#journal    FEBS Lett. (1995) 370:269-272
#title      Identification of a member of mouse semaphorin family.
#cross-references MUID:95385809
#accession  S66498
#status     preliminary
#molecule_type mRNA
#residues   1-834 #label INA
#cross-references EMBL:S79463; NID:g110598; PID:g110599
FEATURE     1-21      #domain signal sequence #status predicted #label SIG\
22-834      #product M-sema F protein #status predicted #label MAT
SUMMARY     #length 834 #molecular-weight 92556 #checksum 7189

Query Match 73.0%; Score 54; DB 2; Length 834;
Best Local Similarity 60.0%; Pred. No. 1.48e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 255 ARVARVCKGD 264
QY 228 SRVAQLCRG 237
|||||

RESULT      10
ENTRY      T00364      #type fragment
TITLE      hypothetical protein KIAA0673 - human (fragment)
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
ACCESSIONS T00364
REFERENCE   Z14142
#authors    Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka,
            A.; Kotani, H.; Nomura, N.; Ohara, O.
#journal    DNA Res. (1998) 5:169-176
#title      Prediction of the coding sequences of unidentified human
            genes. X. The complete sequences of 100 new cDNA clones
            from brain which can code for large proteins in vitro.
#accession  T00364
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues   1-1215 #label ISH
#cross-references EMBL:AB014573; NID:d1204328; PID:d1032609
#experimental_source brain; clone HK02385
GENETICS    KIAA0673
#note       #length 1215 #checksum 8542
SUMMARY

Query Match 71.6%; Score 53; DB 2; Length 1215;
Best Local Similarity 88.9%; Pred. No. 2.34e+00;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 440 SRVAODCRG 448
QY 228 SRVAQLCRG 236
|||||

RESULT      11
ENTRY      A35694      #type complete
TITLE      cut1 protein - fission yeast (Schizosaccharomyces pombe)
ORGANISM    #formal_name Schizosaccharomyces pombe
DATE        30-Nov-1990 #sequence_revision 30-Nov-1990 #text_change
ACCESSIONS A35694
REFERENCE   A35694
#authors    Uzawa, S.; Samejima, I.; Hirano, T.; Tanaka, K.; Yanagida, M.
#journal    Cell (1990) 62:913-925
#title      The fission yeast cut1+ gene regulates spindle pole body

```

```

duplication and has homology to the budding yeast ESp1
gene.
#cross-references MUID:90367123
#accession A35694
##status preliminary
##molecule_type DNA
##residues 1-1827 ##label UZA
#cross-references EMBL:M36179
nucleus
#length 1827 #molecular-weight 209491 #checksum 4643

Query Match 71.6%; Score 53; DB 2; Length 1827;
Best Local Similarity 50.0%; Pred. No. 2.34e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 324 SKVOLCKSE 333
I:| | | | |
Qy 228 SRVQLCRGD 237

RESULT 12
ENTRY C70327 #type complete
TITLE serine--trna ligase (EC 6.1.1.11) - Aquifex aeolicus
ALTERNATE_NAMES seryl-trna synthetase
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
24-Oct-1998
C70327
A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;
Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MUID:98196666
#accession C70327
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-425 ##label AOF
#cross-references GB:AE000682; NID:g2982979; PID:g2982982; GB:AE000657
#experimental_source strain VF5

GENETICS
#gene sers
#description charges tRNA(Ser) with serine
#pathway protein biosynthesis
CLASSIFICATION #superfamily serine--trna ligase
KEYWORDS aminoacyl-trna synthetase; ATP; ligase; protein biosynthesis
SUMMARY #length 425 #molecular-weight 49432 #checksum 5890

Query Match 68.9%; Score 51; DB 2; Length 425;
Best Local Similarity 66.7%; Pred. No. 5.80e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 318 RVVELCTGD 326
I:| | | | |
Qy 229 RVAQLCRGD 237

RESULT 13
ENTRY S75313 #type complete
TITLE serine--trna ligase (EC 6.1.1.11) - Synechocystis sp. (PCC
6803)
ALTERNATE_NAMES protein slr1703; seryl-trna synthetase
ORGANISM #formal_name Synechocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
24-Oct-1998
A75313
S74322
#accessions S75313
REFERENCE S74322

```

```

#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synechocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S75313
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-430 ##label KAN
#cross-references EMBL:D90904; GB:AB001339; NID:g1652225; PID:d1017960;
PID:g1652304
##note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996

GENETICS
#gene sers
#start_codon GTG
FUNCTION
#description charges tRNA(Ser) with serine
#pathway protein biosynthesis
CLASSIFICATION #superfamily serine--trna ligase
KEYWORDS aminoacyl-trna synthetase; ATP; ligase; protein biosynthesis
SUMMARY #length 430 #molecular-weight 48037 #checksum 2062

Query Match 68.9%; Score 51; DB 2; Length 430;
Best Local Similarity 66.7%; Pred. No. 5.80e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 325 RVVELCTGD 333
I:| | | | |
Qy 229 RVAQLCRGD 237

RESULT 14
ENTRY E71017 #type complete
TITLE probable NADH-ubiquinone oxidoreductase subunit - Pyrococcus
horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
14-Aug-1998
A71017
A71000
#accessions A71017
REFERENCE Kwarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#journal DNA Res. (1998) 5:55-76
#title Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references MUID:98344137
#accession E71017
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-173 ##label KAW
#cross-references GB:AP000006; NID:g3236133; PID:d1031484; PID:g3257858
#experimental_source strain OT3
##note this accession replaces an interim accession for a
sequence replaced by GenBank

GENETICS
#gene PH1434
SUMMARY #length 173 #molecular-weight 18895 #checksum 5467

```

```

Query Match          67.6%; Score 50; DB 2; Length 173;
Best Local Similarity 71.4%; Pred. No. 9.05e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 21 RIAQLCK 27
Qy 229 RVAQLCR 235

RESULT 15
ENTRY S43135 #type complete
TITLE serine--trna ligase (EC 6.1.1.11) - Coxiella burnetii
ORGANISM #formal_name Coxiella burnetii
DATE 25-Dec-1994 #sequence_revision 10-Nov-1995 #text_change
24-Oct-1998
ACCESSIONS S43135
REFERENCE S43131
#authors Oswald, W.
#submission submitted to the EMBL Data Library, November 1993
#accession S43135
#molecule_type DNA
#residues 1-423 #label OSW
##cross-references EMBL:X75627; NID:g468527; PID:g468532
##experimental_source isolate nine mile phase I (Bratislava)
GENETICS
#gene sers
FUNCTION
#description charges tRNA(Ser) with serine
#pathway protein biosynthesis
CLASSIFICATION #superfamily serine--trna ligase
KEYWORDS aminoacyl-trna synthetase; ATP; ligase; protein biosynthesis
SUMMARY #length 423 #molecular_weight 48134 #checksum 477

Query Match          67.6%; Score 50; DB 2; Length 423;
Best Local Similarity 66.7%; Pred. No. 9.05e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 318 RVVLCAGD 326
Qy 229 RVAQLCRGD 237

```

Search completed: Thu Jul 8 18:36:26 1999
Job time : 11 secs.

THIS PAGE BLANK (USPTO)


```

RESULT 2
ID VP40_VZVD STANDARD; PRT: 605 AA.
AC P09286;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CAPSID PROTEIN P40 (VIRION STRUCTURAL GENE 33 PROTEIN) [CONTAINS:
GN CAPSID PROTEIN VP24 (PROTEASE) (EC 3.4.21.-); CAPSID PROTEIN VP22A).
OS VARICELLA-ZOSTER VIRUS (STRAIN DUMAS) (VZV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAPERESVIRINAE; VARICELLOVIRUS.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 86306657.
RX DAVISON A.J., SCOTT J.E.;
RT "The complete DNA sequence of varicella-zoster virus.";
RL GEN. VIROL. 67:1759-1816(1986).
CC -!- FUNCTION: VP22A IS A COMPONENT OF THE CAPSID CORE INVOLVED IN
CC PROCESSING AND PACKAGING OF PROGENY DNA. VP24 IS A PROTEASE
CC WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF AND VP22A AT THE C-
CC TERMINUS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: PREFERENTIALLY CLEAVES AT ALA-|-SER OR ALA-|-
CC ALA.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21 (SERINE PROTEASE).
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04370; G60022; -.
DR PIR: G27214; WZBE33.
DR PFAM: PF00716; Peptidase_S21; 1.
DR HSP: P16753; 1CMV.
KW COAT PROTEIN; HYDROLASE; SERINE PROTEASE.
FT CHAIN 1 236 COAT PROTEIN VP24 (PROTEASE).
FT CHAIN 237 7578 COAT PROTEIN VP22A.
FT PROPEP 7579 605 C-TERMINAL PEPTIDE.
FT SITE 236 237 CLEAVAGE (BY THE PROTEASE) (PROBABLE).
FT SITE 578 579 CLEAVAGE (BY THE PROTEASE) (PROBABLE).
FT ACT_SITE 52 52 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 139 139 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 605 AA; 66046 MW; 3CD6BE6B CRC32;

Query Match 75.7%; Score 56; DB 1; Length 605;
Best Local Similarity 66.7%; Pred. No. 2.06e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 293 KVTQLCRND 301
QY 229 RVAQLCRGD 237
!:::!!:::

RESULT 3
ID CUTL1_SCHPO STANDARD; PRT: 1828 AA.
AC P18286;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CUTL1 PROTEIN.
DE CUTL1.
GN CUTL1.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETEALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=972;
RX MEDLINE: 90367123.
RA UZAWA S., SAMEIIMA I., HIRANO T., TANAKA K., YANAGIDA M.;
RT "The fission yeast cutl+ gene regulates spindle pole body duplication
RL and has homology to the budding yeast ESPI gene.";
RN CELL 62:913-925(1990).
[2]
RN REVISIONS.
RP STRAIN=972;
RA FUNABIKI H., KUMADA K., YANAGIDA M.;
RT "Fission yeast Cutl and Cut2 essential for sister chromatid separation
RT concentrate along the metaphase spindle and form large complex.";
RT SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: ESSENTIAL FOR CHROMOSOME DISJUNCTION, REGULATES SPINDLE
CC POLE BODY DUPLICATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: TO YEAST ESPI AND A. NIDULANS BIMB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M36179; G1491780; -.
DR PIR: A35694; A35694.
KW ATP-BINDING; NUCLEAR PROTEIN; CELL DIVISION.
FT NP_BIND 617 624 ATP (POTENTIAL).
SQ SEQUENCE 1828 AA; 209524 MW; A57C6A8F CRC32;

Query Match 71.6%; Score 53; DB 1; Length 1828;
Best Local Similarity 50.0%; Pred. No. 9.42e-01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 324 SKVVLCKSE 333
QY 228 SRVAQLCRGD 237
!:::!!:::

RESULT 4
ID KPVRCANFA STANDARD; PRT: 519 AA.
AC Q29536;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PYRUVATE KINASE, ISOZYME R (EC 2.7.1.40) (FRAGMENT).
GN PKLR.
OS CANIS FAMILIARIS (DOG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=RETICULOCYTES;
RX MEDLINE: 94341326.
RA WHITNEY K.M., GOODMAN S.A., BAILEY E.M., LOTHROP C.D. JR.;
RT "The molecular basis of canine pyruvate kinase deficiency.";
RL EXP. HEMATOL. 22:866-874(1994).
CC -!- CATALYTIC ACTIVITY: ATP + PYRUVATE -> ADP + PHOSPHOENOLPYRUVATE.
CC -!- COFACTOR: REQUIRES MAGNESIUM AND POTASSIUM (BY SIMILARITY).
CC -!- PATHWAY: FINAL STEP IN GLYCOLYSIS.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- DISEASE: DEFECTS IN PKLR ARE A CAUSE INHERITED HEMOLYTIC ANEMIA,
CC AN AUTOSOMAL RECESSIVE DISEASE OF THE BASENJI DOG THAT CLOSELY
CC RESEMBLES HUMAN PK DEFICIENCY.
CC -!- SIMILARITY: BELONGS TO THE PYRUVATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: S71529; G562172; .
 CC PROSITE: PS00110; PYRUVATE_KINASE; 1.
 CC PFAM: PF00224; PK; 1.
 CC HSSP: P11974; LPKN.
 CC TRANSFERASE; KINASE; GLYCOLYSIS; MULTIGENE FAMILY; PHOSPHORYLATION;
 CC MAGNESIUM.
 CC FT NON_TER 1
 CC FT ACT_SITE 258 258 BY SIMILARITY.
 CC FT NON_TER 519 519
 CC FT METAL 260 260 MAGNESIUM (POTENTIAL).
 CC FT METAL 281 281 MAGNESIUM (POTENTIAL).
 CC FT METAL 282 282 MAGNESIUM (POTENTIAL).
 CC SQ SEQUENCE 519 AA; 56251 MW; 5D79D4CE CRC32;

Query Match 70.3%; Score 52; DB 1; Length 519;
 Best Local Similarity 66.7%; Pred. No. 1.52e+00;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 448 AROAHLRCRG 456

QY 228 SRVAQLCRG 236

RESULT 5
 ID SYS_AQUAE STANDARD; PRT; 425 AA.
 AC O66647;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE SERYL-TRNA SYNTHETASE (EC 6.1.1.11) (SERINE--TRNA LIGASE) (SERRS).
 GN SERS.
 OS AQUIFEX AEOLICUS.
 OC BACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE; 98196666.
 RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AGUAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus.";
 RL NATURE 392:353-358(1998).
 CC -!- CATALYTIC ACTIVITY: ATP + L-SERINE + TRNA(SER) = AMP +
 CC -!- PYROPHOSPHATE + L-SERYL-TRNA(SER).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: A600682; G2982982; .
 CC PROSITE: PS00179; AA_TRNA_LIGASE_II_1; 1.
 CC PROSITE: PS00339; AA_TRNA_LIGASE_II_2; FALSE NEG.
 CC AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 CC SEQUENCE 425 AA; 49432 MW; B501ED92 CRC32;

Query Match 68.9%; Score 51; DB 1; Length 425;
 Best Local Similarity 66.7%; Pred. No. 2.52e+00;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 318 RVVELCTGD 326

||: ||| ||

QY 229 RVAQLCRGD 237

RESULT 6
 ID SYS_SYNT3 STANDARD; PRT; 430 AA.
 AC P73201;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE SERYL-TRNA SYNTHETASE (EC 6.1.1.11) (SERINE--TRNA LIGASE) (SERRS).
 GN SERS OR SLR1703.

OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
 OC BACTERIA; CYANOBACTERIA; CHROOCOCCALES; SYNECHOCYSTIS.

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97061201.

RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K.,
 RA OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,
 RA YAMADA M., YASUDA M., TABATA S.;

*Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA RES 3:109-136(1996).

CC -!- CATALYTIC ACTIVITY: ATP + L-SERINE + TRNA(SER) = AMP +
 CC -!- PYROPHOSPHATE + L-SERYL-TRNA(SER).

CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: D90904; G1652304; .

DR PROSITE: PS00179; AA_TRNA_LIGASE_II_1; 1.

DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; FALSE NEG.

DR PFAM: PF00587; TRNA-Synt_2b; 1.

DR HSSP: P34945; LSER.

KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 SQ SEQUENCE 430 AA; 48038 MW; 08BF6135 CRC32;

Query Match 68.9%; Score 51; DB 1; Length 430;
 Best Local Similarity 66.7%; Pred. No. 2.52e+00;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 325 RVVELCTGD 333

||: ||| ||

QY 229 RVAQLCRGD 237

RESULT 7
 ID SYS_COXBU STANDARD; PRT; 423 AA.
 AC P39919;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE SERYL-TRNA SYNTHETASE (EC 6.1.1.11) (SERINE--TRNA LIGASE) (SERRS).

GN SERS

OS COXIELLA BURNETII.

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; COXIELLA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NINE MILE PHASE I / BRATISLAVA;

RA OSWALD W.

RL THESIS (1994), JUSTUS LIEBIG UNIVERSITAET / FRANKFURT, GERMANY.

CC -!- CATALYTIC ACTIVITY: ATP + L-SERINE + TRNA(SER) = AMP +

CC PYROPHOSPHATE + L-SERYL-TRNA(SER).

```
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X75627; G468532; -.
CC PROSITE: PS00179; AA_TRNA_LIGASE_II_1; 1.
CC PROSITE: PS00339; AA_TRNA_LIGASE_II_2; FALSE_NEG.
CC PFAM: PF00587; trna-synt_2b; 1.
CC HSP: P34945; 1SER.
CC AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
CC SEQUENCE 423 AA; 48134 MW; 777FB68A CRC32;
CC -----
CC Query Match 67.6%; Score 50; DB 1; Length 423;
CC Best Local Similarity 66.7%; Pred. No. 4.08e+00;
CC Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC -----
Db 318 RVVELCAGD 326
CC || ||| ||
CC 229 RVAQLCRGD 237
CC -----
RESULT 8
ID TRF1_SALSA STANDARD; PRT; 690 AA.
AC P80426;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE SEROTRANSFERRIN I PRECURSOR (SIDEROPHILIN I) (STF I).
GN STF1.
OS SALMO SALAR (ATLANTIC SALMON).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; PROTACANTHPTERYGII; SALMONIFORMES;
OC SALMONIDAE; SALMO.
[1]
RN RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 94122797.
RA KVINDEDAL A.M., ROERVIK K.A., ALESTROEM P.;
RT "Cloning and characterization of Atlantic salmon (Salmo salar) serum
RT transferrin cDNA.";
RL MOL. MAR. BIOL. BIOTECHNOL. 2:233-238(1993).
[2]
RN RP SEQUENCE OF 1-64 FROM N.A.
RX MEDLINE: 95121925.
RA KVINDEDAL A.M.;
RT "Characterization of the 5' region of the Atlantic salmon (Salmo
RT salar) transferrin-encoding gene.";
RL GENE 150:335-339(1994).
[3]
RN RP SEQUENCE OF 19-37.
RC TISSUE=SERUM;
RA ROED K.H., DEHLI A.K., FLENGSRUD R., MIDTHJELL L., ROERVIK K.A.;
RT "Immunoassay and partial characterization of serum transferrin from
RT Atlantic salmon (Salmo salar L.).";
RL FISH SHELLFISH IMMUNOL. 5:71-80(1995).
CC -!- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE. IT IS RESPONSIBLE FOR THE
CC TRANSPORT OF IRON FROM SITES OF ABSORPTION AND HEME DEGRADATION
CC TO THOSE OF STORAGE AND UTILIZATION. SERUM TRANSFERRIN MAY ALSO
CC HAVE A FURTHER ROLE IN STIMULATING CELL PROLIFERATION.
CC -!- SUBUNIT: MONOMER.
CC -!- TISSUE SPECIFICITY: ABUNDANT IN LIVER AND SERUM WITH SMALLER
CC AMOUNTS FOUND IN THE STOMACH AND KIDNEY.
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
```

```
CC -!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L20313; G431610; -.
CC EMBL: L26909; G598396; -.
CC PROSITE: PS00205; TRANSFERRIN_1; 2.
CC PROSITE: PS00206; TRANSFERRIN_2; 2.
CC PROSITE: PS00207; TRANSFERRIN_3; 1.
CC PFAM: PF00405; transferrin; 2.
CC HSP: P56410; 1AOV.
CC IRON TRANSPORT; GLYCOPROTEIN; METAL-BINDING; MULTIGENE FAMILY; SIGNAL.
CC FT SIGNAL 1 18
CC FT CHAIN 19 690
CC FT DISULFID 28 50
CC FT DISULFID 127 207
CC FT DISULFID 172 186
CC FT DISULFID 235 249
CC FT DISULFID 343 379
CC FT DISULFID 353 370
CC FT DISULFID 404 681
CC FT DISULFID 419 642
CC FT DISULFID 451 529
CC FT DISULFID 475 670
CC FT DISULFID 485 499
CC FT DISULFID 496 512
CC FT DISULFID 569 583
CC FT METAL 74 74
CC FT METAL 104 104
CC FT METAL 201 201
CC FT METAL 257 257
CC FT METAL 394 394
CC FT METAL 428 428
CC FT METAL 523 523
CC FT METAL 591 591
CC FT CARBOHYD 169 169
CC FT BINDING 457 457
CC FT CONFLICT 34 34
CC SEQUENCE 690 AA; 74597 MW; 1C3D942E CRC32;
CC -----
CC Query Match 67.6%; Score 50; DB 1; Length 690;
CC Best Local Similarity 60.0%; Pred. No. 4.08e+00;
CC Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
CC -----
Db 180 SKLCLCKGD 189
CC I::|::|:|
CC 228 SRVAQLCRGD 237
CC -----
RESULT 9
ID TRF2_SALSA STANDARD; PRT; 691 AA.
AC P80429;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE SEROTRANSFERRIN II PRECURSOR (SIDEROPHILIN II) (STF II).
GN STF2.
OS SALMO SALAR (ATLANTIC SALMON).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; PROTACANTHPTERYGII; SALMONIFORMES;
OC SALMONIDAE; SALMO.
[1]
RN RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 94122797.
RA KVINDEDAL A.M., ROERVIK K.A., ALESTROEM P.;
RT "Cloning and characterization of Atlantic salmon (Salmo salar) serum
```


transferrin cDNA.";
RL MOL. MAR. BIOL. BIOTECHNOL. 2:233-238(1993).
[2]
RN SEQUENCE OF 19-37.
RP TISSUE-SERUM;
RA ROED K.H., DEHLI A.K., FLENGSRUD R., MIDTHJELL L., ROERVIK K.A.;
RC "Immunosay and partial characterization of serum transferrin from
RT Atlantic salmon (Salmo salar L.).";
RL FISH SHELLFISH IMMUNOL. 5:71-80(1995).
CC -!- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE. IT IS RESPONSIBLE FOR THE
CC TRANSPORT OF IRON FROM SITES OF ABSORPTION AND HEME DEGRADATION
CC TO THOSE OF STORAGE AND UTILIZATION. SERUM TRANSFERRIN MAY ALSO
CC HAVE A FURTHER ROLE IN STIMULATING CELL PROLIFERATION.
CC
CC -!- SUBUNIT: MONOMER.
CC
CC -!- TISSUE SPECIFICITY: ABUNDANT IN LIVER AND SERUM WITH SMALLER
CC AMOUNTS FOUND IN THE STOMACH AND KIDNEY.
CC
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC
CC -!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
DR PROSITE: PS00205; TRANSFERRIN_1; 2.
DR PROSITE: PS00206; TRANSFERRIN_2; 2.
DR PROSITE: PS00207; TRANSFERRIN_3; 1.
DR PFAM: PF00405; transferrin; 2.
DR HSP: P56410; LAOV.
KW IRON TRANSPORT; GLYCOPROTEIN; METAL-BINDING; MULTIGENE FAMILY; SIGNAL.
FT SIGNAL 1 18
FT CHAIN 19 691 SEROTRANSFERRIN 2.
FT DISULFID 28 50 BY SIMILARITY.
FT DISULFID 127 207 BY SIMILARITY.
FT DISULFID 172 186 BY SIMILARITY.
FT DISULFID 235 249 BY SIMILARITY.
FT DISULFID 343 379 BY SIMILARITY.
FT DISULFID 353 370 BY SIMILARITY.
FT DISULFID 404 682 BY SIMILARITY.
FT DISULFID 419 643 BY SIMILARITY.
FT DISULFID 451 530 BY SIMILARITY.
FT DISULFID 475 671 BY SIMILARITY.
FT DISULFID 485 499 BY SIMILARITY.
FT DISULFID 496 513 BY SIMILARITY.
FT DISULFID 570 584 BY SIMILARITY.
FT METAL 74 74 IRON 1 (BY SIMILARITY).
FT METAL 104 104 IRON 1 (BY SIMILARITY).
FT METAL 201 201 IRON 1 (BY SIMILARITY).
FT METAL 257 257 IRON 1 (BY SIMILARITY).
FT METAL 394 394 IRON 2 (BY SIMILARITY).
FT METAL 428 428 IRON 2 (BY SIMILARITY).
FT METAL 524 524 IRON 2 (BY SIMILARITY).
FT METAL 592 592 IRON 2 (BY SIMILARITY).
FT CARBOHYD 169 169 POTENTIAL.
FT BINDING 457 457 ANION (POTENTIAL).
SQ SEQUENCE 691 AA; 74663 MW; 95893BE6 CRC32;

Query Match 67.6%; Score 50; DB 1; Length 691;
Best Local Similarity 60.0%; Pred. No. 4.08e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 180 SKLCLCKGKD 189
:: |||||
QY 228 SRVAQLCRGD 237

RESULT 10
ID Y07N-MYCTU STANDARD; PRT; 250 AA.
AC Q11029;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 26.9 KD PROTEIN CY02B10.23.
GN MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERIAE; MYCOBACTERIAE; MYCOBACTERIUM.

[1]
RN SEQUENCE FROM N.A.
RP STRAIN-H37RV;
RA MURPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: SOME, TO CLASS-3 OF ADENYLYL CYCLASES.
CC -!- SIMILARITY: TO M.TUBERCULOSIS MTCV31.19C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 275555; E250359; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 250 AA; 26886 MW; 8F8F508C CRC32;

Query Match 66.2%; Score 49; DB 1; Length 250;
Best Local Similarity 55.6%; Pred. No. 6.57e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 195 RIAQLCHPN 203
:: ||||| :
QY 229 RVAQLCRGD 237

RESULT 11
ID POLL-SCICO STANDARD; PRT; 1004 AA.
AC Q03277;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RETROVIRUS-RELATED POL. POLYPROTEIN FROM TYPE I RETROTRANSPOSABLE
DE ELEMENT R1 [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE] (FRAGMENT).
OS SCIARA COPROPHILA (FUNGUS GNAT).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; NEMATOCERA; SCIAROIDAE; BRADYSIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93196484.
RA BURKE W.D., EICKBUSH D.G., XIONG Y., JAKUBCZAK J.L., EICKBUSH T.H.;
RT "Sequence relationship of retrotransposable elements R1 and R2 within
RT and between divergent insect species".
RL MOL. BIOL. EVOL. 10:163-185(1993).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
CC N PYROPHOSPHATE + DNA(N).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00945; G160852; -.
DR PIR; A44490; A44490.
DR PFAM; PF00078; rvc; 1.
KW TRANSFERASE; RNA-DIRECTED DNA POLYMERASE; TRANSPOSABLE ELEMENT;
KW HYDROLASE; NUCLEASE; ENDONUCLEASE.
FT NON_TER 1 1
FT DOMAIN <1 852 REVERSE TRANSCRIPTASE.
FT DOMAIN 853 1004 NUCLEIC ACID-BINDING ENDONUCLEASE.
SQ SEQUENCE 1004 AA; 113433 MW; F7FC591A CRC32;

Query Match 66.2%; Score 49; DB 1; Length 1004;
Best Local Similarity 66.7%; Pred. No. 6.57e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Db 341 GRVFKLCRG 349
QY 228 SRVAQLCRG 236

RESULT 12
ID NEST_HUMAN STANDARD; PRT: 1618 AA.
AC P48681; O00552;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NESTIN.
GN NES.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE:PLACENTA;
RX MEDLINE; 93123384.
RA DAHLSTRAND J., MCKAY R.D.G., ZIMMERMAN L.B., LENDAHN U.;
RT "Characterization of the human nestin gene reveals a close
RT evolutionary relationship to neurofilaments.";
RL J. CELL SCI. 103:589-597(1992).
RN [2]
RP SEQUENCE OF 296-309 FROM N.A.
RA YAMORSKY P.J., KAPPEN C.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- TISSUE SPECIFICITY: CNS STEM CELLS.
CC -!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
CC IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X65964; G35019; -
DR EMBL; AF004335; G2209204; -
DR MIM; 600915; -
DR PROSITE; PS00226; IF: 1.
KW INTERMEDIATE FILAMENT; COILED COIL; HEPTAD REPEAT PATTERN;
KW NEURONE.
FT DOMAIN 1 7 HEAD.
FT DOMAIN 8 312 ROD.
FT DOMAIN 313 1618 TAIL.
FT DOMAIN 8 43 COIL 1A.
FT DOMAIN 44 55 LINKER 1.
FT DOMAIN 56 151 COIL 1B.
FT DOMAIN 152 172 LINKER 12.
FT DOMAIN 173 191 COIL 2A.
FT DOMAIN 192 194 LINKER 2.
FT DOMAIN 195 312 COIL 2A.
SQ SEQUENCE 1618 AA; 176704 MW; C0572CAD CRC32;

Query Match 66.2%; Score 49; DB 1; Length 1618;
Best Local Similarity 60.0%; Pred. No. 6.57e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 486 SRVFSTCRGE 495
QY 228 SRVAQLCRGD 237

RESULT 13
ID YOAM_BACSU STANDARD; PRT: 313 AA.
AC P45910;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
HYPOTHETICAL 36.1 KD PROTEIN IN SPOIIIC-CWLA INTERGENIC REGION.
GN YOAM.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / JH642;
RX MEDLINE; 95219086.
RA TAKEMARU K.-I., MIZUNO M., SATO T., TAKEUCHI M., KOBAYASHI Y.;
RT "Complete nucleotide sequence of a skin element excised by DNA
RT rearrangement during sporulation in Bacillus subtilis.";
RL MICROBIOLOGY 141:323-327(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / JH642;
RA KOBAYASHI Y., MIZUNO M., MASUDA S., TAKEMARU K., HOSONO S., SATO T.,
RA TAKEUCHI M.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP IDENTIFICATION.
RX MEDLINE; 96084975.
RA MEDIGUE C., MOSZER I., VIARI A., DANCHIN A.;
RT "Analysis of a Bacillus subtilis genome fragment using a co-operative
RT computer system prototype.";
RL GENE 165:GC37-GC51(1995).
CC -!- SIMILARITY: TO B.SUBTILIS YQXC AND T.HYODYSENTERIAE HEMOLYSIN
CC TLYA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D32216; G1217842; -
DR EMBL; D84432; D1013067; -
DR EMBL; Z99117; E1183855; -
DR SUBTILIS; BG11264; YOAM.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 313 AA; 36138 MW; 4AD21054 CRC32;

Query Match 64.9%; Score 48; DB 1; Length 313;
Best Local Similarity 62.5%; Pred. No. 1.05e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 285 SRIHQMCRCR 292
QY 228 SRVAQLCR 235

RESULT 14
ID VINT_BPHKO STANDARD; PRT: 357 AA.
AC P16407;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-MAY-1991 (REL. 18, LAST ANNOTATION UPDATE)
DE INTEGRASE.
GN INT.
OS BACTERIOPHAGE HK022.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE;
OC LAMBDA PHAGE GROUP.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89342457.
RA YAGIL E., DOLEV S., OBERTO J., KISLEV N., RAMAIAH N.,
RA WEISBERG R.A.;
RT "Determinants of site-specific recombination in the lambdaoid
RT coliphage HK022. An evolutionary change in specificity.";
RL J. MOL. BIOL. 207:695-717(1989).
```

CC -!- FUNCTION: INTEGRASE IS NECESSARY FOR INTEGRATION OF THE PHAGE INTO
CC THE HOST GENOME BY SITE-SPECIFIC RECOMBINATION. IN CONJUNCTION
CC WITH EXCISIONASE, INTEGRASE IS ALSO NECESSARY FOR EXCISION OF THE
CC PROPHAGE FROM THE HOST GENOME.
CC
CC -!- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X51962; GI5760; -
CC PIR: S04990; S04990.
CC PFAM: PF00589; Phage_integrase; 1.
CC HSP: P03700; IAE9.
CC DNA RECOMBINATION; DNA INTEGRATION.
CC ACT_SITE 342 PROBABLE TRANSIENT COVALENT LINKAGE TO
CC SEQUENCE 357 AA; 40358 MW; B38FD29F CRC32;
CC
CC Query Match 64.9%; Score 48; DB 1; Length 357;
CC Best Local Similarity 71.4%; Pred. No. 1.05e+01;
CC Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
CC
CC Db 212 RVGLCLR 218
CC ||::|||
CC QY 229 RVAQLCR 235
CC
CC RESULT 15
CC ID YP33_YEAST STANDARD; PRT; 446 AA.
CC AC P19541;
CC DT 01-FEB-1991 (REL. 17, CREATED)
CC DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
CC DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
CC DE PUTATIVE TRANSCRIPTIONAL REGULATORY PROTEIN IN MKK2-COX11 INTERGENIC
CC REGION.
CC GN YPL133C OR LP12C.
CC OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
CC OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
CC OC SACCHAROMYCETACEAE; SACCHAROMYCES.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RA HALL J., DEPAULO T., AHMED A., BUSSEY H., FORTIN N., FRIESEN J.D.,
CC RA STORMS R.K., VO D.H., WANG Y., WINNETT E.;
CC RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC RN [2]
CC RP SEQUENCE OF 1-95 FROM N.A.
CC RX MEDLINE; 90360986.
CC RA TRAGOLOFF A., CAPITANIO N., NOBREGA M.P., GATTI D.;
CC RT "Cytochrome oxidase assembly in yeast requires the product of COX11,
CC RT a homolog of the P. denitrificans protein encoded by ORF3.";
CC RL EMBO J. 9:2759-2764(1990).
CC RN [3]
CC RP PRESENCE OF A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR CLUSTER.
CC RX MEDLINE; 93284106.
CC RA BORK P., OUZOUNIS C., SANDER C., SCHARF M., SCHNEIDER R.,
CC RA SONNHAMMER E.;
CC RT "Comprehensive sequence analysis of the 182 predicted open reading
CC RT frames of yeast chromosome III.";
CC RL PROTEIN SCI. 1:1677-1690(1992).
CC CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U43703; GI244781; -
CC EMBL: X55731; G3566; -
CC PIR: S12795; S12795.
CC DR PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
CC DR PROSITE; PS00048; ZN2_CY6_FUNGAL_2; 1.
CC DR PFAM; PF00172; Zn_c1us; 1.
CC KW HYPOTHETICAL PROTEIN; TRANSCRIPTION REGULATION; DNA-BINDING;
CC KW NUCLEAR PROTEIN; ZINC; METAL-BINDING.
CC FT DNA_BIND 15 45 ZN(2)-CYS(6), FUNGAL-TYPE.
CC SQ SEQUENCE 446 AA; 50081 MW; 64DD5143 CRC32;
CC
CC Query Match 64.9%; Score 48; DB 1; Length 446;
CC Best Local Similarity 62.5%; Pred. No. 1.05e+01;
CC Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
CC
CC Db 41 IAHLCRED 48
CC QY 230 VAQLCRGD 237
CC
CC Search completed: Thu Jul 8 18:35:26 1999
CC Job time : 8 secs.

THIS PAGE BLANK (USPTO)

Result No.	Query %			ID	Description	Pred. No.	
	Score	Match	Length				
1	74	100.0	393	11	098371	SEMAPHORIN L (FRAGMENT	3.75e+04
2	74	100.0	666	40	075326	SEMAPHORIN L.	3.75e+04
3	65	87.8	900	10	035004	UNCONVENTIONAL MYOSIN.	3.09e+02
4	63	85.1	1126	10	081840	MYOSIN HEAVY CHAIN - L	7.97e-02
5	58	78.4	748	11	062177	SEMAPHORIN A PRECURSOR	7.94e-01
6	58	78.4	749	4	013214	SEMAPHORIN V.	7.94e-01
7	58	78.4	750	4	093018	SEMAPHORIN V.	7.94e-01
8	56	75.7	161	4	043561	ACYL CARRIER PROTEIN,	1.94e+00
9	56	75.7	782	11	062179	SEMAPHORIN C (SEM C) (1.94e+00
10	55	74.3	205	2	032716	MDCG.	3.00e+00
11	55	74.3	205	2	P71426	MDCF.	3.00e+00
12	55	74.3	354	2	P74242	HYPOTHETICAL 39.3 KD P	3.00e+00
13	55	74.3	760	11	062178	SEMAPHORIN B PRECURSOR	3.00e+00
14	54	73.0	834	11	064151	SEMAPHORIN I (M-SEMA F	4.64e+00
15	53	71.6	828	10	082630	STELAR K+ OUTWARD RECT	7.13e+00
16	53	71.6	861	11	009126	SEMAPHORIN J (SEMAPHOR	7.13e+00
17	53	71.6	1215	4	075161	KIAA0873 PROTEIN (FRAG	7.13e+00
18	52	70.3	888	11	035464	SEMAPHORIN VIA.	1.09e+01
19	51	68.9	862	4	092854	SEMAPHORIN.	1.66e+01
20	51	68.9	1023	4	092502	MYELOBLAST KIAA0189.	1.66e+01

OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GORN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030698; G3523115; -
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 74; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 3.75e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 260 SRVAQLCRGD 269
QY 228 SRVAQLCRGD 237

RESULT 3 PRELIMINARY; PRT; 900 AA.
ID O50004;
AC O50004;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE UNCONVENTIONAL MYOSIN.
GN HMY3.
OS HELIANTHUS ANNUUS (COMMON SUNFLOWER).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; ASTERALES; ASTERACEAE; HELIANTHUS.
RN [1]
RP SEQUENCE FROM N.A.
RA VUGREK O.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U94783; G2731702; -
SQ SEQUENCE 900 AA; 101216 MW; 96D16DB9 CRC32;

Query Match 87.8%; Score 65; DB 10; Length 900;
Best Local Similarity 70.0%; Pred. No. 3.09e-02;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 338 SRVQICRGE 347
QY 228 SRVAQLCRGD 237

RESULT 4 PRELIMINARY; PRT; 1126 AA.
ID O81840;
AC O81840;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE MYOSIN HEAVY CHAIN - LIKE PROTEIN.
GN M122.180.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RA REICHERT B.J., BAREL E., HOEISEL J., MEWES H.W., MAYER K.,
RA SCHUELLER C., BEVAN M.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AL030978; E1308257; -
SQ SEQUENCE 1126 AA; 128758 MW; E9C84E22 CRC32;

Query Match 85.1%; Score 63; DB 10; Length 1126;
Best Local Similarity 80.0%; Pred. No. 7.97e-02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 329 SRVAQLCNGE 338
QY 228 SRVAQLCRGD 237

RESULT 5 PRELIMINARY; PRT; 748 AA.
ID O62177;
AC O62177;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN A PRECURSOR (SEM A).
GN SEMAA OR SEMA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC LOW LEVELS FOUND BETWEEN DAYS 10-12.
CC EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL
BIRTH.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85990; G854324; -
DR MGD; MGI:107561; SEMAA.
DR PFAM; PF00047; Ig; 1.
KW SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KW DEVELOPMENTAL PROTEIN.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 748 SEMAPHORIN A.
FT DOMAIN 586 649 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 748 AA; 82894 MW; A7E53A8D CRC32;

Query Match 78.4%; Score 58; DB 11; Length 748;
Best Local Similarity 70.0%; Pred. No. 7.94e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 262 SRVQICRND 271
QY 228 SRVAQLCRGD 237

RESULT 6 PRELIMINARY; PRT; 749 AA.
ID Q13214;
AC Q13214;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96210603.
RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
cancer deletion region and demonstrate distinct expression
patterns.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR EMBL; U28369; G974284; -
DR PFAM; PF00047; Ig; 1.

```

SQ SEQUENCE 749 AA; 83121 MW; 6138C5D5 CRC32;
  Query Match 78.4%; Score 58; DB 4; Length 749;
  Best Local Similarity 70.0%; Pred. No. 7.94e-01;
  Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
  Db 263 SRVQICRND 272
  QY 228 SRVAQLCRGD 237
  III::I::I

RESULT 7
ID Q93018 PRELIMINARY; PRT; 750 AA.
AC Q93018;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS SEMAPHORIN (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA DANTE M., WANSLEY P.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U73167; G2860035; -.
DR PFAM; PF00047; ig; 1.
SQ SEQUENCE 750 AA; 83034 MW; CBC87E34 CRC32;
  Query Match 78.4%; Score 58; DB 4; Length 750;
  Best Local Similarity 70.0%; Pred. No. 7.94e-01;
  Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
  Db 264 SRVQICRND 273
  QY 228 SRVAQLCRGD 237
  III::I::I

RESULT 8
ID O14561 PRELIMINARY; PRT; 161 AA.
AC O14561;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ACYL CARRIER PROTEIN, (ACP)
DE (NADH-UBIQUINONE OXIDOREDUCTASE 9.6 KD SUBUNIT) (5' PARTIAL)
DE (FRAGMENT).
GN A-735G6.1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA ADAMS M.D., LOFTUS B.J., ZHOU L., CROSBY M., FUHRMANN J., BRANDON R.,
RA KIM U.J., KERLAVAGE A.R., VENTER J.C.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AC002400; G2576345; -.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; 1.
DR PFAM; PF00550; pp-binding; 1.
KW UBIQUINONE.
FT NON_TER 1
SQ SEQUENCE 161 AA; 17930 MW; 29B859B4 CRC32;
  Query Match 75.7%; Score 56; DB 4; Length 161;
  Best Local Similarity 75.0%; Pred. No. 1.94e+00;
  Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
  Db 265 SRVQICRND 272
  QY 228 SRVAQLCRGD 237
  III::I::I

RESULT 9
ID Q62179 PRELIMINARY; PRT; 782 AA.
AC Q62179;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN C (SEM C) (FRAGMENT).
GN SEMAC OR SEMC.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI; TISSUE-BRAIN;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
RL NEURON 14:941-948(1995).
CC -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC -!- SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC -!- LOW LEVELS FOUND BETWEEN DAYS 10-12.
CC EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL
CC BIRTH.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85992; G854328; -.
DR MGI; MGI:107559; SEMAC.
KW IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KW DEVELOPMENTAL PROTEIN.
FT NON_TER 1
FT DOMAIN 550 607 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 782 AA; 86823 MW; CCB0E95D CRC32;
  Query Match 75.7%; Score 56; DB 11; Length 782;
  Best Local Similarity 70.0%; Pred. No. 1.94e+00;
  Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
  Db 226 SRVARVCKGD 235
  QY 228 SRVAQLCRGD 237
  III::I::I::I

RESULT 10
ID Q32716 PRELIMINARY; PRT; 205 AA.
AC Q32716;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MDCG.
GN MDCG.
OS KLEBSIELLA PNEUMONIAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC KLEBSIELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97352552.
RA HOENKE S., SCHMID M., DIMROTH P.;
RT "Sequence of a gene cluster from Klebsiella pneumoniae encoding
RT malonate decarboxylase and expression of the enzyme in Escherichia
RT coli."
RL EUR. J. BIOCHEM. 246:530-538(1997).
DR EMBL; U95087; G2240017; -.
SQ SEQUENCE 205 AA; 23014 MW; ADC0E9F6 CRC32;
```

Query Match 74.3%; Score 55; DB 2; Length 205;
Best Local Similarity 72.7%; Pred. No. 3.00e+00;
Matches 8; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

Db 155 TRVAQLPCRAD 165
Qy 228 SRVAQL-CRGD 237

RESULT 11
ID P71426 PRELIMINARY; PRT; 205 AA.
AC P71426;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE MDCF.
OS KLEBSIELLA PNEUMONIAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC KLEBSIELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CG43;
RA CHANG H., DENG W., CHAOU S., LEE R., PENG H.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U56096; G1314844; -;
SQ SEQUENCE 205 AA; 23023 MW; EB79C797 CRC32;

Query Match 74.3%; Score 55; DB 2; Length 205;
Best Local Similarity 72.7%; Pred. No. 3.00e+00;
Matches 8; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

Db 155 TRVAQLPCRAD 165
Qy 228 SRVAQL-CRGD 237

RESULT 12
ID P74242 PRELIMINARY; PRT; 354 AA.
AC P74242;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 39.3 KD PROTEIN.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGITURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. PCC6803. II. Sequence determination of the entire
genome and assignment of potential protein-coding regions."
RL DNA RES. 3:109-136(1996).
DR EMBL; D90913; D1019069; -;
DR PFAM; PF00534; Glycos_transf_1; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 354 AA; 39290 MW; 0CFBA22F CRC32;

Query Match 74.3%; Score 55; DB 2; Length 354;
Best Local Similarity 77.8%; Pred. No. 3.00e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 137 NRVAQLCPG 145
Qy 228 SRVAQLCRG 236

RESULT 13
ID Q62178 PRELIMINARY; PRT; 760 AA.
AC Q62178;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN B PRECURSOR (SEM B).
GN SEMAB OR SEMB.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI; TISSUE=BRAIN;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC LOW LEVELS FOUND BETWEEN DAYS 10-12.
CC EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL
BIRTH.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85991; G854326; -;
DR MGD; MGI:107560; SEMAB.
KW SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KW DEVELOPMENTAL PROTEIN.
FT SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 760 SEMAPHORIN B.
FT DOMAIN 574 629 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 760 AA; 83458 MW; D3990F12 CRC32;

Query Match 74.3%; Score 55; DB 11; Length 760;
Best Local Similarity 70.0%; Pred. No. 3.00e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 263 SRVAQVCKND 272
Qy 228 SRVAQLCRGD 237

RESULT 14
ID Q64151 PRELIMINARY; PRT; 834 AA.
AC Q64151;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN I (M-SEMA F-A FACTOR IN NEURAL NETWORK DEVELOPMENT).
GN SEMAI OR M-SEMA F.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95385809.
RA INAGAKI S., FURUYAMA T., IWAHASHI Y.;
RT "Identification of a member of mouse semaphorin family.";
RL FEBS LETT. 370:269-272(1995).
DR EMBL; S79463; G110599; -;
DR MGD; MGI:109252; SEMAI.
SQ SEQUENCE 834 AA; 92556 MW; A3565ACF CRC32;

Query Match 73.0%; Score 54; DB 11; Length 834;
Best Local Similarity 60.0%; Pred.No. 4.64e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 255 ARVARYCKGD 264
QY 228 SRVAQLCRGD 237
:||||:|

RESULT 15
ID O82630 PRELIMINARY; PRT; 828 AA.
AC O82630;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DE STELAR K+ OUTWARD RECTIFYING CHANNEL.
GN SKOR.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG ERECTA;
RX MEDLINE: 98412660.
RA GAYMARD F., PILOT G., LACOMBE B., BOUCHEZ D., BRUNEAU D.,
RA BOUCHEREZ J., MICHAUX-FERRIERE N., THIBAUD J.B., SENTENAC H.;
RT "Identification and disruption of a plant shaker-like outward channel
involved in K+ release into the xylem sap.";
RL CELL 94:647-655(1998).
DR EMBL: AJ223358; E1321636; -;
SQ SEQUENCE 828 AA; 93898 MW; AA41E3DD CRC32;

Query Match 71.6%; Score 53; DB 10; Length 828;
Best Local Similarity 85.7%; Pred.No. 7.13e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 487 RVAELCR 493
QY 229 RVAQLCR 235
|||||

Search completed: Thu Jul 8 18:35:57 1999
Job time : 12 secs.

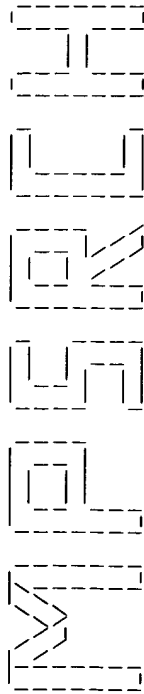
THIS PAGE BLANK (USPTO)

CC inhibiting activity. The semaphorin gene is useful for the diagnosis,

DR WPI; 98-312416/27.

by sequence, and

Sequence



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:38:51 1999; Maspar time 8.52 Seconds
Tabular output not generated. 512.773 Million cell updates/sec

Title: >US-09-041-236-2
Description: (230-338) from US09041236.pap (16 of 45)
Perfect Score: 811
Sequence: 1 VAQLCRGDDGGESSVSWSK.....LKGYSLLPNRPGRKCLPDQ 109

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 4008593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 39.757; Variance 71.771; scale 0.554

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	499	61.5	653	2	T03102 semaphorin homolog A3	4.65e-84
2	281	34.6	749	2	G01856 semaphorin V - human	7.90e-37
3	277	34.2	751	2	I48748 semaphorin E - mouse	5.37e-36
4	272	33.5	748	2	I48744 semaphorin A - mouse	5.86e-35
5	266	32.8	772	2	I48747 semaphorin D - mouse	1.02e-33
6	264	32.6	772	2	A90699 collapsin - chicken	2.64e-33
7	263	32.4	403	2	E42521 A39R protein - vacci	4.25e-33
8	258	31.8	782	2	I48746 semaphorin C - mouse	4.54e-32
9	257	31.7	771	2	D49423 semaphorin III precu	7.29e-32
10	256	31.6	441	2	S29921 hypothetical protein	1.17e-31
11	256	31.6	666	2	I58169 semaphorin III - mous	1.17e-31
12	255	31.4	834	2	S6498 M-sema F protein prec	1.88e-31
13	233	28.7	753	2	G02173 semaphorin III family	5.61e-27
14	224	27.6	760	2	I48745 semaphorin B - mouse	3.61e-25
15	160	19.7	656	2	B49423 semaphorin I - fruit	8.28e-13
16	152	18.7	1074	2	JC5928 semaphorin F precurs	2.41e-11
17	151	18.6	711	2	A49423 semaphorin I precurs	3.65e-11
18	149	18.4	724	2	C49423 semaphorin II precurs	8.39e-11
19	135	16.6	730	2	JH0798 fasciclin IV precurs	2.55e-08
20	133	16.4	295	2	JQ1775 SalL9R protein - vacc	5.67e-08
21	110	13.6	122	2	J01845 14R protein - variola	3.88e-04
22	110	13.6	122	2	H36852 A43R protein - variol	3.88e-04
23	108	13.3	142	2	JQ1776 SalFLR protein - vacc	8.06e-04

24 103 12.7 406 2 B69064 conserved hypotheticala 4.86e-03
25 94 11.6 947 1 P8BY2P H+-transporting ATPas 1.10e-01
26 92 11.3 261 2 T00106 ABC-type transport pr 2.14e-01
27 91 11.2 498 1 HYBSPA pseudolysin (EC 3.4.2 2.98e-01
28 88 10.9 211 2 S77532 rehydrin - Synecocys 7.94e-01
29 88 10.9 918 1 P8BY1P H+-transporting ATPas 7.94e-01
30 87 10.7 445 2 B70146 glycine-tRNA ligase 1.10e+00
31 87 10.7 456 2 JQ2396 helper component - zu 1.10e+00
32 87 10.7 456 2 JQ2397 helper component - zu 1.10e+00
33 87 10.7 456 2 JQ2395 helper component - zu 1.10e+00
34 87 10.7 715 2 B71418 hypothetical protein 1.10e+00
35 84 10.4 213 2 S29910 hypothetical protein 2.83e+00
36 84 10.4 305 2 F70381 glycyl-tRNA synthetas 2.83e+00
37 84 10.4 307 2 A36953 yscQ protein - yersin 2.83e+00
38 84 10.4 459 2 D70914 hypothetical protein 2.83e+00
39 84 10.4 563 2 A37184 glucan-binding protei 2.83e+00
40 84 10.4 895 1 PXCKP H+-transporting ATPas 2.83e+00
41 83 10.2 457 2 S56459 hypothetical 48.5K pr 3.86e+00
42 83 10.2 498 2 C36851 A29L protein - variol 3.86e+00
43 83 10.2 2514 1 MNWVW2 nonstructural polypro 3.86e+00
44 82 10.1 288 2 S64243 hypothetical protein 5.26e+00
45 82 10.1 1058 2 S08436 pol polyprotein - hum 5.26e+00

ALIGNMENTS

RESULT 1
ENTRY T03102 #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
T03102
24-Mar-1999
ACCESSIONS T03102
REFERENCE Z14840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type DNA
#residues 1-653 #label ENS
#cross-references EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY #length 653 #molecular-weight 73645 #checksum 5501

Query Match 61.5%; Score 499; DB 2; Length 653;
Best Local Similarity 56.0%; Pred. No. 4.65e-84;
Matches 61; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

Db 284 VGVQSSDQGGESSLSVYKTTFLKARLACVDYDTGTGRIYNELQDIFIWQAPENSWEETLI 343
Qy 230 VAQLCRGDDGGESSLSVSKWNTFLKAMLCVSDAATNKNRLQDVFLPLDPDSGQWRDTRV 289
Db 344 YGLFLSPWNTSAVCVFVKDIDHVEKTSKLNKYNHKLPTPRPGCKMKNH 392
Qy 290 YGVFSPNPNYSAVCVYSLGIDIDKVFRTSSLSKGYHSSLPNRPGRKCLPDQ 338

RESULT 2
ENTRY G01856 #type complete
TITLE semaphorin V - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
04-Sep-1998
ACCESSIONS G01856
REFERENCE G08634
#authors Sekido, Y.
#submission submitted to the EMBL Data Library, June 1995
#accession G01856
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues 1-749 #label SEK
#cross-references EMBL:U28369; NID:g974283; PID:g974284
CLASSIFICATION #superfamily semaphorin


```

RESULT      9
ENTRY       D49423          #type complete
TITLE       semaphorin III precursor - human
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
           04-Sep-1998
ACCESSIONS  D49423
REFERENCE   A49423
#authors    Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
#journal     Cell (1993) 75:1389-1399
#title       The Semaphorin genes encode a family of transmembrane a
            secreted growth cone guidance molecules.
#accession   D49423
##status     preliminary; nucleic acid sequence not shown
##molecule_type mRNA
##residues   1-771 ##label KOL
##cross-references GB:L26081; NID:g799328; PID:g436560
GENETICS
#gene        GDB:SEMA1
##cross-references GDB:283448
CLASSIFICATION #superfamily semaphorin
SUMMARY      #length 771 #molecular-weight 88889 #checksum 6249

Query Match      31.7%; Score 257; DB 2; Length 771;
Best Local Similarity 44.0%; Pred. No. 7.29e-32;
Matches 40; Conservative 23; Mismatches 19; Indels 9; Gaps

Db 265 IGOICKNDGGHRSI-VNKKWTFILKARLICSVPGPNGITDHFDELDQVFLM-NFKDP-KN 3
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 230 VAQLCRGDGGESSLSVSKWNTEFLKAMLVCS-DA-ATNKNFNLQDVFLLPDPSPGQWRD 2
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 322 PVYGVGTFTSSNIIFKGSAVCYMYSMSDVRRVF 352
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 287 TRYGVFSNPWN-Y-SAVCVISLGDIDKVF 314

RESULT      10
ENTRY       S29921          #type complete
TITLE       hypothetical protein 15 - vaccinia virus
ORGANISM    #formal_name vaccinia virus
DATE        20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
           09-Sep-1997
ACCESSIONS  S29921
REFERENCE   S29907
#authors     Amegadzie, B.Y.
#submission submitted to the EMBL Data Library, January 1991
#accession   S29921
##status     preliminary
##molecule_type DNA
##residues   1-441 ##label AME
##cross-references EMBU:X57318; NID:g622239; PID:g62254
SUMMARY      #length 441 #molecular-weight 50185 #checksum 6034

Query Match      31.6%; Score 256; DB 2; Length 441;
Best Local Similarity 41.1%; Pred. No. 1.17e-31;
Matches 44; Conservative 18; Mismatches 39; Indels 6; Gaps

Db 229 IAQMCLNDEGGSSLSHSRWSTFLKVELEC-DID-GRSY-R-Q-IIHSRTIKTD-NDTIL 2
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 230 VAQLCRGDGGESSLSVSKWNTEFLKAMLVCSDAATKNFNELQDVFLLPDPSPGQWRDTRV 2

Db 283 YVFEDSPYSKALCYISMNTWKQSFTSKLEGYTKQLPSPASGICLP 329
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 290 YGVFSNPWNYSACVYSLGIDDKVFRTSSLKGYHSSLPNRPKGICLP 336

RESULT      11
ENTRY       I58169         #type fragment
TITLE       semaphorin III - mouse (fragment)
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change

```

```
04-Sep-1998
ACCESSIONS I58169
REFERENCE I58169
#authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;
#journal Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#title Neuron (1995) 14:949-959
#cross-references MUID:95267432
#accession I58169
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-666 ##label RES
#cross-references GB:L40484; NID:g703190
GENETICS SemaIII
#gene Semaphorin
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 666 #checksum 9654

Query Match 31.6%; Score 256; DB 2; Length 666;
Best Local Similarity 42.9%; Pred. No. 1.17e-31;
Matches 39; Conservative 25; Mismatches 18; Indels 9; Gaps 7;

Db 159 IGQICKNDFGHRSL-VNKWTFELKARLCSVPGPGNGIDTHFDELDQVFLM-NSKDP-KN 215
Qy 230 VAQLCRGDQGGESSLSVKWNTFLKAMLVCS--DAATNKNFNRLQDVFLLPDPGQWRD 286
Db 216 PIVYGVFTTSSNIFKGSVAVCMYSMDVRRVL 246
Qy 287 TRYGVFSNPW-Y--SVCVYSLGDIDKVF 314

RESULT 12
ENTRY S66498 #type complete
TITLE M-sema F protein precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change
10-Sep-1997
ACCESSIONS S66498
REFERENCE S66498
#authors Inagaki, S.; Furuyama, T.; Iwahashi, Y.
#journal FEBS Lett. (1995) 370:269-272
#title Identification of a member of mouse semaphorin family.
#cross-references MUID:95385809
#accession S66498
#status preliminary
#molecule_type mRNA
#residues 1-834 ##label INA
#cross-references EMBL:S79463; NID:g1110598; PID:g1110599
FEATURE
1-21 #domain signal sequence #status predicted #label SIG\
22-834 #product M-sema F protein #status predicted #label MAT
SUMMARY #length 834 #molecular-weight 92556 #checksum 7189

Query Match 31.4%; Score 255; DB 2; Length 834;
Best Local Similarity 43.8%; Pred. No. 1.88e-31;
Matches 42; Conservative 16; Mismatches 31; Indels 7; Gaps 4;

Db 257 VARVCKDGMGARTLQ-KKWTTFELKARLVCSAPDKVYFNOLKAVHLRGAS--WHNTTF 313
Qy 230 VAQLCRGDQGGESSLSVKWNTFLKAMLVCSDAATNKNFNRLQDVFLLPDPGQWRDTRV 289
Db 314 FGVOARGMGMDLSAVCEYQLEQIOQVF-EGPYKEY 348
Qy 290 YGVFSNPW---NYSACVYSLGDIDKVFRTSSLKGY 322

RESULT 13
ENTRY G02173 #type complete
TITLE semaphorin III family homolog - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
04-Sep-1998
```

```
ACCESSIONS G02173
REFERENCE G09275
#authors Naylor, S.
#submissionsubmitted to the EMBL Data Library, October 1995
#accession G02173
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-753 ##label NAY
#cross-references EMBL:U38276; NID:g1061350; PID:g1061351
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 753 #molecular-weight 84941 #checksum 5681

Query Match 28.7%; Score 233; DB 2; Length 753;
Best Local Similarity 41.8%; Pred. No. 5.61e-27;
Matches 38; Conservative 24; Mismatches 20; Indels 9; Gaps 6;

Db 264 IGRICLNDGGHCL-VNKNSTFLKARLVCSVPGEIGIETHFDELDQVVF-QQTQDV-RN 320
Qy 230 VAQLCRGDQGGESSLSVKWNTFLKAMLVCS--DAATNKNFNRLQDVFLLPDPGQWRD 286
Db 321 PVIYAVFTSSGVSFRGSVAVCYSMADIRMVF 351
Qy 287 TRYGVFSNP-WNY--SACVYSLGDIDKVF 314

RESULT 14
ENTRY I48745 #type complete
TITLE semaphorin B - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
28-Feb-1997
ACCESSIONS I48745
REFERENCE I48744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Murine semaphorin D/collapsin is a member of a diverse gene
#title family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48745
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-760 ##label RES
#cross-references EMBL:X85991; NID:g854325; PID:g854326
GENETICS
#gene semB
SUMMARY #length 760 #molecular-weight 83458 #checksum 2188

Query Match 27.6%; Score 224; DB 2; Length 760;
Best Local Similarity 42.9%; Pred. No. 3.61e-25;
Matches 39; Conservative 18; Mismatches 25; Indels 9; Gaps 5;

Db 265 VAQVCKNDVGEKLLQ-KKWTTFELKARLVCSAPDKVYFNOLKAVHLRGAS--P-SVSRI 320
Qy 230 VAQLCRGDQGGESSLSVKWNTFLKAMLVCSDAATNKNFNRLQDVFLLPDPGQWRDTRV 289
Db 321 YAVFTSQWVGGRSSAVCAFSLTDIRVEFK 351
Qy 290 YGVFSNPWY-----SACVYSLGDIDKVF 315

RESULT 15
ENTRY B49423 #type fragment
TITLE semaphorin I - fruit fly (Drosophila melanogaster) (fragment)
ORGANISM #formal_name Drosophila melanogaster
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
16-Feb-1997
ACCESSIONS B49423
REFERENCE A49423
#authors Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
#journal Cell (1993) 75:1389-1399
#title The Semaphorin genes encode a family of transmembrane and
#accession B49423 secreted growth cone guidance molecules.
```

```
##status      preliminary: nucleic acid sequence not shown; not
               compared with conceptual translation
##molecule_type mRNA
##residues    1-656 #label KOL
##cross-references GB:L26082

GENETICS
#gene          semai
##cross-references FlyBase:FBgn0011259
SUMMARY      #length 656 #checksum 3860

Query Match  19.7%, Score 160; DB 2; Length 656;
Best Local Similarity 35.6%; Pred. No. 8.28e-13;
Matches 32; Conservative 16; Mismatches 36; Indels 6; Gaps 5;

Db 137 VARVCKWDKGGPHRF-NRWTSELKSLNCSIPGDYFFYENEIQASNLVEGQYGSMSK 195
QY 230 VAOLCRGDQGGESSLSVSKWNTFLKAMLVCSDAATNK-NFNRLQDVF-LLPDPSGQWRDT 287

Db 196 LIYGVNTPSNIPGSVACAFALQDIADTF 225
QY 288 RYGVFSNPWN-Y--SACVYSLGDIDKVF 314
```

Search completed: Thu Jul 8 18:39:02 1999
Job time : 11 secs.

THIS PAGE BLANK (USPTO)

W P E R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:37:44 1999; MasPar time 4.98 Seconds
Tabular output not generated. 618.816 Million cell updates/sec

Title: >US-09-041-236-2
Description: (230-338) from US09041236.pap (16 of 45)
Perfect Score: 811
Sequence: 1 VAQLCRGQGGESSLSVSKW.....LKGYHSSLPNRPKGKCLPDQ 109

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 40.747; Variance 63.822; scale 0.638

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.
1	263	32.4	403	1 VA39_VACCC PROTEIN A39.	3.06e-38
2	256	31.6	441	1 VA39_VACCCV PROTEIN A39.	1.37e-36
3	99	12.2	916	1 PNA1_AJECV PLASMA MEMBRANE ATPASE	3.08e-03
4	94	11.6	947	1 PNA2_YEAST PLASMA MEMBRANE ATPASE	2.12e-02
5	91	11.2	498	1 ELAS_PSEAE PSEUDOLYSIN PRECURSOR	6.52e-02
6	88	10.9	918	1 PNA1_YEAST PLASMA MEMBRANE ATPASE	1.96e-01
7	87	10.7	3083	1 POLG_ZYMYR GENOME POLYPROTEIN [CO	2.81e-01
8	85	10.5	444	1 CHLB_CHLPT PROTOCHLOROPHYLLIDE RE	5.74e-01
9	85	10.5	444	1 PNA1_KLULA PLASMA MEMBRANE ATPASE	5.74e-01
10	84	10.4	213	1 AT12_VACCV PUTATIVE A-TYPE INCLUS	8.15e-01
11	84	10.4	307	1 YSCQ_YERPE YOP PROTEINS TRANSLOCA	8.15e-01
12	84	10.4	307	1 YSCQ_YERPE YOP PROTEINS TRANSLOCA	8.15e-01
13	84	10.4	895	1 PNA1_CANAL PLASMA MEMBRANE ATPASE	8.15e-01
14	83	10.2	457	1 MPL_ECCLI UDP-N-ACETYLMURAMATE:L	1.16e+00
15	83	10.2	2514	1 POLN_ONNVG NONSTRUCTURAL POLYPROT	1.16e+00
16	83	10.2	3080	1 POLG_ZYMYC GENOME POLYPROTEIN [CO	1.16e+00
17	82	10.1	288	1 NTF3_YEAST NGG1-INTERACTING FACTO	1.63e+00
18	82	10.1	436	1 THCB_RHOSO CYTOCHROME P450 116 (E	1.63e+00
19	82	10.1	1058	1 POL_RV2D2 POL POLYPROTEIN [CONTA	1.63e+00
20	81	10.0	334	1 DCAM_BOVIN S-ADENOSYLMETHIONINE D	2.29e+00
21	81	10.0	334	1 DCAM_HUMAN S-ADENOSYLMETHIONINE D	2.29e+00
22	81	10.0	684	1 NU5C_SESIN NADH-PLASTOQUINONE OXI	2.29e+00
23	81	10.0	2291	1 RRPB_BEV RNA-DIRECTED RNA POLYM	2.29e+00

24	81	10.0	2431	1 POLN_SFV NONSTRUCTURAL POLYPROT	2.29e+00
25	80	9.9	198	1 CLYT_CLYGR CLYTIN PRECURSOR (PHIA	3.22e+00
26	80	9.9	431	1 MRKA_TRYBB PUTATIVE SERINE/THREON	3.22e+00
27	80	9.9	828	1 MRKC_KLEPN OUTER MEMBRANE USHER P	3.22e+00
28	79	9.7	216	1 INAA_ECCLI INAA PROTEIN.	4.50e+00
29	79	9.7	238	1 Y4TF_RHISN PROBABLE AMINO-ACID AB	4.50e+00
30	79	9.7	299	1 Y025_MYCPN HYPOTHETICAL PROTEIN M	4.50e+00
31	79	9.7	322	1 AT12_VACCC PUTATIVE A-TYPE INCLUS	4.50e+00
32	79	9.7	575	1 VG05_BP14 TAIL-ASSOCIATED LYSOZY	4.50e+00
33	79	9.7	1131	1 YANC_SCHPO HYPOTHETICAL 128.8 KD	4.50e+00
34	79	9.7	1666	1 CLH_SCHPO PROBABLE CLATHRIN HEAV	4.50e+00
35	79	9.7	1895	1 YLK3_CABEL PUTATIVE SERINE/THREON	4.50e+00
36	78	9.6	333	1 DCAM_RAT S-ADENOSYLMETHIONINE D	6.26e+00
37	78	9.6	334	1 DCAM_MESAU S-ADENOSYLMETHIONINE D	6.26e+00
38	78	9.6	334	1 DCAM_MOUSE S-ADENOSYLMETHIONINE D	6.26e+00
39	78	9.6	360	1 HR38_BOMMO PROBABLE NUCLEAR HORMO	6.26e+00
40	78	9.6	414	1 SPY1_RAT SERINE--PYRUVATE AMINO	6.26e+00
41	78	9.6	499	1 CPN1_MESAU CYTOCHROME P450 XIB1 P	6.26e+00
42	78	9.6	637	1 LBR_CHICK LAMIN B RECEPTOR.	6.26e+00
43	78	9.6	920	1 PMA1_ZIGRO PLASMA MEMBRANE ATPASE	6.26e+00
44	78	9.6	920	1 PMA1_NEUCR PLASMA MEMBRANE ATPASE	6.26e+00
45	77	9.5	380	1 RFC_SHIDY O-ANTIGEN POLYMERASE.	8.68e+00

ALIGNMENTS

RESULT 1	STANDARD;	PRT:	403 AA.
ID	VA39_VACCC		
AC	P21062;		
DT	01-FEB-1991 (REL. 17, CREATED)		
DT	01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)		
DT	01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)		
DE	PROTEIN A39.		
GN	A39R.		
OS	VACCINIA VIRUS (STRAIN COPENHAGEN).		
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;		
QC	ORTHOPOXVIRUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 91021027.		
RA	GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,		
RA	PAOLETTI E.;		
RT	"The complete DNA sequence of vaccinia virus.";		
RL	VIROLOGY 179:247-266(1990).		
RN	[2]		
RP	COMPLETE GENOME.		
RA	GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,		
RA	PAOLETTI E.;		
RL	VIROLOGY 179:517-563(1990).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M35027; G335517; -		
DR	PIR; E42521; E42521. -		
SQ	SEQUENCE 403 AA; 45741 MW; FE0DB9AB CRC32;		
	Query Match 32.4%; Score 263; DB 1; Length 403;		
	Best Local Similarity 42.1%; Pred. No. 3.06e-38;		
	Matches 45; Conservative 17; Mismatches 39; Indels 6; Gaps 6;		

Db	191	IAQMLNDEGGPSLSHRWSTLKEVLEC-DID-GRSY-R-Q-IHSRTIKTD-NDTIL 244
QY	230	VAQLCRGQGGESSLSVSKWNTLKAFLVSDAATNKNFNRLQDFLLPDPSPGQWRDTRV 289
Db	245	YVFEDSPYSKALCYTSMNTIKOSFSTSKLEGYTKQLPSPAGICLP 291
QY	290	IGVFSNPWNYSACVYSLGDIDKVKFRTSSLUKYHSSLNRPNGKCLP 336

```

RESULT      2
ID          VA39_VACCV        STANDARD;           PRT;       441 AA.
P24764;
DT          01-MAR-1992 (REL. 21, CREATED)
DT          01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT          01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE          PROTEIN A39.
GN          A39R OR (SALL9R AND Salfir).
OS          VACCINIA VIRUS (STRAIN WR).
OC          VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OR          ORTHOPOXVIRUS.
[1]
SEQUENCE FROM N.A.
RN          MEDLINE: 91310644.
RA          AMEGADZIE B.Y., AHN B.-Y., MOSS B.:
RT          "Identification, sequence, and expression of the gene encoding a Mr
RL          35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase.";
RJ          J. BIOL. CHEM. 266:13712-13718(1991).
[2]
SEQUENCE FROM N.A.
RN          MEDLINE: 91259063.
RA          SMITH G.L., CHAN Y.S., HOWARD S.T.:
RT          "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RL          the right inverted terminal repeat.";
RJ          J. GEN. VIROL. 72:1349-1376(1991).
CC          -!- CAUTION: IN REF.2 THIS SEQUENCE IS IN TWO PART (SALL9R AND
CC          Salfir) DUE TO THE PRESENCE OF FRAMESHIFT COMPARED TO THAT OF
CC          REF.1.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licensel@isb-sib.ch).
-----
EMBL: M61187; G335798; -
DR          EMBL: D11079; G22726; ALT_SEQ.
DR          EMBL: X57318; G62454; -.
DR          PIR: S29921; S29921.
SQ          SEQUENCE 441 AA; 50185 MW; 2C823A68 CRC32;
Query Match              31.6%   Score 256; DB 1: Length 441;
Best Local Similarity    41.1%; Pred.No. 1.37e-36;
Matches 44; Conservative 18; Mismatches 39; Indels 6; Gaps 6;

Db             229 IAQCLNDGSGPSLSHRWSTFLKVELEC-DID-GRSY-R-Q-IHSRTIKTD-NDTL 282
               :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Qy            230 VAQLCRDGGESSLSYSKWNTFLKMLVCSDAATNKFNRLQDVLLPDSPGWTRTV 289
               |::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

Db             283 YVFEDSPYSKALCTYSMTNIKQSFTSKLEGYTKLPSPASGICLP 329
               ||::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Qy            290 YGVFSPNPWSAVCVYSLGDIDKVFTSSLKGHYSHSLPRPGKCLP 336
               ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

RESULT      3
ID          PMAJ_AJECA        STANDARD;           PRT;       916 AA.
Q07421;
DT          01-FEB-1995 (REL. 31, CREATED)
DT          01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT          01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE          PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).
GN          PMAJ.
OS          AJELLOMYCES CAPSULATA (HISTOPLASMA CAPSULATUM).
OC          EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCAMYCETES; PLECTOMYCETES;
ON          ONYGNALES; ONYGENACEAE; AJELLOMYCES.
[1]
SEQUENCE FROM N.A.
RN          MEDLINE: 94124018.
RA          SCHAFER M.P., DEAN G.E.;
```


CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL: X03534; G4187; ..
 CC DR EMBL: 272530; E243842; ..
 CC DR PIR: A25823; PXBYIP.
 CC DR SGD: L0001449; PMAL.
 CC DR PROSITE: PS00154; ATPASE_E1_E2; 1.
 CC DR PFAM: PF00122; E1-E2-ATPase; 1.
 CC KW HYDROLASE: HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
 CC ATP-BINDING; MULTIGENE FAMILY; GLYCOPROTEIN.
 CC FT DOMAIN 1 115
 CC FT TRANSMEM 116 136 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 137 140 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 141 160
 CC FT DOMAIN 161 291 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 292 313 3 (POTENTIAL).
 CC FT DOMAIN 314 325 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 326 347
 CC FT DOMAIN 348 719 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 720 738 5 (POTENTIAL).
 CC FT DOMAIN 739 754 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 755 774 6 (POTENTIAL).
 CC FT DOMAIN 775 824 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 825 845 7 (POTENTIAL).
 CC FT DOMAIN 846 857 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 858 874 8 (POTENTIAL).
 CC FT DOMAIN 875 918 CYTOPLASMIC (POTENTIAL).
 CC FT MOD_RES 378 378 PHOSPHORYLATION (PROBABLE).
 CC FT BINDING 474 474 ATP (BY SIMILARITY).
 CC FT CARBOHYD 848 848 POTENTIAL.
 CC FT DOMAIN 5 17 POLY-SER.
 CC FT DOMAIN 31 78 ASP/GLU-RICH (ACIDIC).
 CC FT DOMAIN 39 44 POLY-ASP.
 CC FT DOMAIN 585 590 POLY-GLY.
 CC FT SEQUENCE 918 AA: 99619 MW; 93CIFEA3 CRC32;
 SQ

Query Match 10.9%; Score 88; DB 1; Length 918;
 Best Local Similarity 26.0%; Pred. No. 1.96e-01;
 Matches 20; Conservative 19; Mismatches 29; Indels 9; Gaps 9;

Db 766 LATGSWIT-ITWFL-PKGIIIONFGAMGIMFLQISLTENLWIFTPRAAGPFWSSIPSW 823
 QY 244 LSVSKWNTFLKAMLVCSDAATNKNFNRLQDV-FLLPDPGQW-R-DPRVYGVF-SN-P-W 297
 Db 824 QLAGA-VEAVDIIATMF 839
 QY 298 NYSACVVISLGDIDKVF 314

RESULT 7
 ID POLG-ZYMVR STANDARD; PRT; 3083 AA.
 AC Q89330;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DE GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
 DE COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KD PROTEIN
 DE 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (C1); 6 KD PROTEIN 2 (6K2);
 DE GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
 DE (EC 3.4.22.-) (49 KD PROTEINASE) (49 KD-PRO); NUCLEAR INCLUSION
 DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
 DE COAT PROTEIN (CP)].
 OS ZUCCHINI YELLOW MOSAIC VIRUS (STRAIN REUNION ISLAND) (ZYMV).
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
 CC POTYVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95146958.
 RA WISLER G.C., PORCIFULL D.E., HIEBERT E.;

RT "Characterization of the P1 protein and coding region of the zucchini
 RT yellow mosaic virus.";
 RL J. GEN. VIROL. 76:37-45(1995).
 CC -!- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHD
 CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
 CC -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
 CC MAY BE INVOLVED IN REPLICATION
 CC -!- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
 CC -!- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
 CC -!- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
 CC POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
 CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
 CC INDIVIDUAL PROTEINS.
 CC -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
 CC -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
 CC -!- SIMILARITY: BELONGS TO THE POTYVIRUSES POTYPROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL: L29569; G461421; ..
 CC DR PFAM: PF00271; helicase_C; 1.
 CC DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
 CC DR PFAM: PF00767; Poty-coat; 1.
 CC DR PFAM: PF00851; Peptidase_C6; 1.
 CC DR PFAM: PF00863; Peptidase_C4; 1.
 CC KW HYDROLASE; TRANSFERASE; THIOL PROTEASE; RNA-DIRECTED RNA POLYMERASE;
 CC COAT PROTEIN; POLYPROTEIN; COVALENT PROTEIN-RNA LINKAGE; HELICASE;
 CC ATP-BINDING.
 CC FT CHAIN 1 307 N-TERMINAL PROTEIN.
 CC FT CHAIN 308 769 HELPER COMPONENT PROTEINASE.
 CC FT CHAIN 770 2 PROTEIN P3
 CC FT CHAIN 1167 1167 6 KD PROTEIN 1.
 CC FT CHAIN 1168 1801 CYTOPLASMIC INCLUSION PROTEIN.
 CC FT CHAIN 1802 1854 6 KD PROTEIN 2.
 CC FT CHAIN 1855 2044 GENOME-LINKED PROTEIN.
 CC FT CHAIN 2045 2287 NUCLEAR INCLUSION PROTEIN A.
 CC FT CHAIN 2288 2804 NUCLEAR INCLUSION PROTEIN B.
 CC FT CHAIN 2805 3083 COAT PROTEIN.
 CC FT NP_BIND 1252 1259 ATP (POTENTIAL).
 CC SQ SEQUENCE 3083 AA; 351158 MW; D541F982 CRC32;

Query Match 10.7%; Score 87; DB 1; Length 3083;
 Best Local Similarity 44.4%; Pred. No. 2.81e-01;
 Matches 16; Conservative 6; Mismatches 10; Indels 4; Gaps 4;

Db 658 NIFL-AMLVNNEAKDFTKMIRDV-LIPML-GQW 690
 QY 250 NTFKAMLVCSDAATNKNFNRL-QDVFLLPDPGQW 284

RESULT 8
 ID CHLB_CHLPT STANDARD; PRT; 444 AA.
 AC P37824;
 DT 01-OCT-1994 (REL. 30, CREATED)
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
 DE PROTOCHLOROPHYLLIDE REDUCTASE CHLB SUBUNIT (EC 1.3.1.33) (NADPH-
 DE PROTOCHLOROPHYLLIDE OXIDOREDUCTASE CHLB SUBUNIT) (FRAGMENT).
 GN CHLB.
 OS CHLAMYDOMONAS PITSCHMANNII.
 OC CHLOROPLAST.
 OC EUKARYOTA; VIRIDIPANTAE; CHLOROPHYTA; CHLOROPHYCEAE; VOLVOCALES;
 CC CHLAMYDOMONADACEAE; CHLAMYDOMONAS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94033309.
 RA LIU X.-Q., XU H., HUANG C.;

RT "Chloroplast chlB gene is required for light-independent chlorophyll
RT accumulation in Chlamydomonas reinhardtii.";
RL PLANT MOL. BIOL. 23:297-308(1993).
CC -!- FUNCTION: INVOLVED IN THE LIGHT-INDEPENDENT ACCUMULATION OF
CC CHLOROPHYLL, PROBABLY AT THE STEP OF REDUCTION OF PROTO-
CC CHLOROPHYLLIDE TO CHLOROPHYLLIDE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: CHLOROPHYLLIDE A + NADP(+) "
CC PROTOCHLOROPHYLLIDE + NADPH.
CC -!- PATHWAY: LIGHT-INDEPENDENT CHLOROPHYLL BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE CHLB/BCHB FAMILY.
DR PFAM: PF00148; oxidored.nitro. 1.
DR MENDEL: 16353; Chlpt:chlB:mn16353.
KW CHLOROPLAST; PHOTOSYNTHESIS; CHLOROPHYLL BIOSYNTHESIS; OXIDOREDUCTASE;
KW NADP.
FT NON_TER 444 444
SQ SEQUENCE 444 AA; 50058 MW; B9C005C5 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 444;
Best Local Similarity 50.0%; Pred. No. 5.74e-01;
Matches 10; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 270 EIEKILNLSNGYQSSLP 289
: : : : :
QY 309 DIDKVFSTSLKGYHSLPN 328
: : : : :

RESULT 9
ID PMAL_KLULA STANDARD; PRT; 899 AA.
AC P49380;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).
GN PMAL.
OS KLUYVEROMYCES LACTIS (YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; KLUYVEROMYCES.
RN [1]
SEQUENCE FROM N.A.
RC STRAIN:WM37;
RX MEDLINE: 95247668.
RA MIRANDA M., RAMIREZ J., PENA A., CORIA R.;
RT "Molecular cloning of the plasma membrane H(+)-ATPase from
RT Kluyveromyces lactis: a single nucleotide substitution in the gene
RT confers ethidium bromide resistance and deficiency in K+ uptake.";
RL J. BACTERIOL. 177:2360-2367(1995).
CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING
CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKINIZATION MAY MEDIATE
CC GROWTH RESPONSES.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -!- ENZYME REGULATION: ACTIVATED BY HIGH PH OR ALSO BY POTASSIUM IONS
CC WHEN THE MEDIUM PH IS LOW.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L37875; G598435; -.
DR PROSITE: PS00154; ATPASE_E1_E2; 1.
DR PFAM: PF00122; E1-E2_ATPase; 1.
KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
KW ATP-BINDING. 1 96 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 97 117
TRANSMEM 97 117

FT DOMAIN 118 121 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 122 141 2 (POTENTIAL).
FT DOMAIN 142 272 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 273 294 3 (POTENTIAL).
FT DOMAIN 295 305 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 306 328 4 (POTENTIAL).
FT DOMAIN 329 700 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 701 719 5 (POTENTIAL).
FT DOMAIN 720 735 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 736 755 6 (POTENTIAL).
FT DOMAIN 756 805 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 806 826 7 (POTENTIAL).
FT DOMAIN 827 838 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 839 855 8 (POTENTIAL).
FT DOMAIN 856 899 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 359 359 PHOSPHORYLATION (PROBABLE).
FT BINDING 455 455 ATP (BY SIMILARITY).
FT DOMAIN 566 571 POLY-GLY.
FT VARIANT 569 669 M -> I (IN MUTANT 3.3; LOW CAPACITY TO
FT PUMP OUT PROTONS).
SQ SEQUENCE 899 AA; 98259 MW; 1FC57DAE CRC32;

Query Match 10.5%; Score 85; DB 1; Length 899;
Best Local Similarity 29.9%; Pred. No. 5.74e-01;
Matches 23; Conservative 15; Mismatches 30; Indels 9; Gaps 9;

Db 747 LAIGTWIT-LTTMFV-PKGGIIONFGSDGVLFLOITSENWLIIFITRAAGPFWSSIPSW 804
: : : : :
QY 244 LSVKWNFTFLKAMLYCSDAATNKNFNRLQDV-FLLPDSGQW-R-DTRVYGVF-SN-P-W 297
: : : : :

Db 805 QLSGA-VLIVDIIATMF 820
: : : : :
QY 298 NYSACVYSLSGDIKVF 314
: : : : :

RESULT 10
ID AT12_VACCV STANDARD; PRT; 213 AA.
AC P24758;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE PUTATIVE A-TYPE INCLUSION PROTEIN (PROTEIN A26).
OS VACCINIA VIRUS (STRAIN WR).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE: 91310644.
RA AMEGADZIE B.Y., AHN B.-Y., MOSS B.;
RT "Identification, sequence, and expression of the gene encoding a Mr
RT 35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase.";
RL J. BIOL. CHEM. 266:13712-13718(1991).
CC -!- FUNCTION: MATURE VIRUSES ARE OCCLUDED INTO THE ATI. AND IT HAS
CC BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING
CC DISSEMINATION FROM ANIMAL TO ANIMAL.
CC -!- A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF LARGE MASSES
CC WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF INFECTED CELLS.
CC -!- SIMILARITY: TO COWPOX VIRUS A-TYPE INCLUSION PROTEIN C-TERMINUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61187; G335786; -.
DR EMBL: X57318; G62243; -.
DR PIR: S29910; S29910.
SQ SEQUENCE 213 AA; 25037 MW; EBD1F80B CRC32;

Query Match 10.4%; Score 84; DB 1; Length 213;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M74075; G170818; -
CC PIR; A41336; PXCKP
CC -----

DR PROSITE: PS00154; ATPASE_E1_E2; 1.
DR PFAM: PF00122; E1-E2-ATPase; 1.
KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
KW ATP-BINDING; MULTIGENE FAMILY.
FT DOMAIN 1 92 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 93 113 1 (POTENTIAL).
FT DOMAIN 114 117 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 118 137 2 (POTENTIAL).
FT DOMAIN 138 268 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 269 301 3 (POTENTIAL).
FT DOMAIN 291 320 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 302 324 4 (POTENTIAL).
FT DOMAIN 325 596 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 697 715 5 (POTENTIAL).
FT DOMAIN 716 731 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 732 751 6 (POTENTIAL).
FT DOMAIN 752 801 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 802 822 7 (POTENTIAL).
FT DOMAIN 823 834 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 835 851 8 (POTENTIAL).
FT DOMAIN 852 895 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 355 355 PHOSPHORYLATION (PROBABLE).
FT BINDING 451 451 ATP (BY SIMILARITY).
SQ SEQUENCE 895 AA: 97459 MW: 97459 MW: AFF22471 CRC32;

Query Match 10.4%; Score 84; DB 1; Length 895;
Best Local Similarity 28.6%; Pred. No. 8.15e-01;
Matches 12; Conservative 11; Mismatches 16; Indels 3; Gaps 3;

Db 743 LAIGTWT-LTMM-L-PKGIIQNFGLGILFLQISLTENW 782

QY 244 LSVSKWTFKAMLVGSDAATNKNFNLDQV-FLPDPGQW 284

RESULT 14

ID MPL-ECOLI STANDARD; PRT; 457 AA.
AC P37773; P76804;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE UDP-N-ACETYLMURAMATE-L-ALANYL-GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE
DE LIGASE (EC 6.3.2.-).
GN MPL.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN:K12 / MG1655;
RX MEDLINE: 95334362.
RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
RA BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RL NUCLEIC ACIDS RES. 23:2105-2119(1995).
RN [2]
RN SEQUENCE OF 1-64 FROM N.A.
RX MEDLINE: 86335617.
RA HAMILTON W.D.O., HARRISON D.A., DYER T.A.;
RT "Sequence of the Escherichia coli fructose-1,6-bisphosphatase gene."
RL NUCLEIC ACIDS RES. 16:8707-8707(1988).
RN [3]
RN IDENTIFICATION.
RX MEDLINE: 95075659.
RA BORODOVSKY M., RUDD K.E., KOONIN E.V.;
RT "Intrinsic and extrinsic approaches for detecting genes in a
RT bacterial genome."
RL NUCLEIC ACIDS RES. 22:4756-4767(1994).
RN [4]
RN CHARACTERIZATION.
RX MEDLINE: 96404780.
RA MENGIN-LECREULX D., VAN HEIJENOORT J., PARK J.T.;
RT "Identification of the mpl gene encoding UDP-N-acetylmuramate:

RT L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase in Escherichia
RT coli and its role in recycling of cell wall peptidoglycan.";
RL J. BACTERIOL. 178:5347-5352(1996).
CC -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. REUTILIZES THE INTACT
CC TRIPEPTIDE L-ALANYL-GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE.
CC -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14003; G537075; -.
DR EMBL: AE000494; G1790680; -.
DR EMBL: X12545; -; NOT_ANNOTATED_CDS.
DR ECOGENE; EGI2440; MPL.
DR HSSP; P14900; IUAG.
KW PEPTIDOGLYCAN SYNTHESIS; CELL DIVISION; CELL WALL; LIGASE;
KW ATP-BINDING.
FT NP_BIND 110 116 ATP (POTENTIAL).
SQ SEQUENCE 457 AA: 49874 MW: A2BF8001 CRC32;

Query Match 10.2%; Score 83; DB 1; Length 457;
Best Local Similarity 29.0%; Pred. No. 1.16e-00;
Matches 18; Conservative 16; Mismatches 23; Indels 5; Gaps 5;

Db 341 LAAL-RGKVGGTARIIVLEPRSNMTK-MGICKDDIAP-SLGRADEVFLQPAHPQVA 397

QY 230 VAQLCRGQDGESSL-SV-SKWNTEFLKAMLVGSDAATNKNFNLDQVFLPDPGQW 287

Db 398 EV 399

QY 288 RV 289

RESULT 15

ID POLN-ONNVG STANDARD; PRT; 2514 AA.
AC P13886;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE NONSTRUCTURAL POLYPROTEIN [CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO
DE NSP4].
OS O'NYONG-NYONG VIRUS (STRAIN GULU) (ONN).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
OC ALPHAVIRUS.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 90177205.
RA LEVINSON R.S., STRAUSS J.H., STRAUSS E.G.;
RT "Complete sequence of the genomic RNA of O'nyong-nyong virus and its
RT use in the construction of alphavirus phylogenetic trees."
RL VIROLOGY 175:110-123(1990).
RN [2]
RN SEQUENCE OF 1334-2514 FROM N.A.
RX MEDLINE: 88206074.
RA STRAUSS E.G., LEVINSON R., RICE C.M., DALRYMPLE J., STRAUSS J.H.;
RT "Nonstructural proteins nsp3 and nsp4 of Ross River and O'Nyong-nyong
RT viruses: sequence and comparison with those of other alphaviruses."
RL VIROLOGY 164:265-274(1988).
CC -!- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```
CC      EMBL; M20303; G332559; -.
DR      PIR; A34680; MNWVN2.
DR      PIR; B28614; MNWVNN.
DR      PFAM; PF00978; RNA_dep_RNAPol12; 2.
KW      POLYPROTEIN; NONSTRUCTURAL PROTEIN; RNA-BINDING; HELICASE.
FT      CHAIN 1 535     NONSTRUCTURAL PROTEIN NSP1.
FT      CHAIN 536 1333   NONSTRUCTURAL PROTEIN NSP2.
FT      CHAIN 1334 1903   NONSTRUCTURAL PROTEIN NSP3.
FT      CHAIN 1904 2514   NONSTRUCTURAL PROTEIN NSP4.
SQ      SEQUENCE 2514 AA; 280114 MW; F63F4F08 CRC32;

Query Match          10.2%; Score 83; DB 1; Length 2514;
Best Local Similarity 34.3%; Pred No. 1,15e+00;
Matches 12; Conservative 10; Mismatches 11; Indels 2; Gaps 2;
```

Search completed: Thu Jul 8 18:37:55 1999
Job time : 11 secs.

THIS PAGE BLANK (USPTO)

W O R L D

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:38:15 1999; MasPar time 12.51 Seconds
Tabular output not generated. 475.518 Million cell updates/sec

Title: >US-09-041-236-2
Description: (230-338) from US09041236.pep (16 of 45)
Perfect Score: 811

Sequence: 1 VAQLCRGDQGGESSLSVSKW.....LKGYSLLNPRPGKCLPDQ 109

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 39.961; Variance 65.646; scale 0.609

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	811	100.0	666	4	SEMAPHORIN L.	6.86e-170
2	769	94.8	393	11	SEMAPHORIN L (FRAGMENT)	2.52e-159
3	499	61.5	653	14	SIMILAR TO GENBANK ACC	2.59e-92
4	281	34.6	749	4	SEMAPHORIN V.	1.81e-40
5	281	34.6	750	4	SEMAPHORIN V.	1.81e-40
6	278	34.3	294	13	COLLAPSPIN-3 (FRAGMENT)	8.76e-40
7	278	34.3	751	13	COLLAPSPIN-3	8.76e-40
8	277	34.2	751	11	SEMAPHORIN E PRECURSOR	1.48e-39
9	272	33.5	748	11	SEMAPHORIN A PRECURSOR	2.04e-38
10	271	33.4	761	13	COLLAPSPIN-2	3.44e-38
11	269	33.2	751	4	SEMAPHORIN E.	9.79e-38
12	266	32.8	772	11	SEMAPHORIN D PRECURSOR	4.69e-37
13	266	32.8	772	11	SEMAPHORIN III/COLLAPS	4.69e-37
14	264	32.6	772	13	COLLAPSPIN.	1.33e-36
15	258	31.8	782	11	SEMAPHORIN C (SEM C) (3.01e-35
16	257	31.7	771	4	SEMAPHORIN-III.	5.06e-35
17	257	31.7	861	11	SEMAPHORIN J (SEMAPHOR	5.06e-35
18	255	31.4	834	11	SEMAPHORIN I (M-SEMA F	1.43e-34
19	233	28.7	295	13	COLLAPSPIN-4 (FRAGMENT)	1.16e-29
20	233	28.7	753	4	SEMAPHORIN III FAMILY	1.16e-29

21	233	28.7	754	11	088633	SEMAPHORIN IV ISOFORM	1.16e-29
22	233	28.7	785	4	Q15704	SEMAPHORIN.	1.16e-29
23	233	28.7	785	4	Q13275	SEMAPHORIN IV.	1.16e-29
24	233	28.7	785	11	088632	SEMAPHORIN IV ISOFORM	1.16e-29
25	227	28.0	775	11	P70275	SEMAPHORIN H.	2.44e-28
26	225	27.7	775	4	Q15041	KIAA0331.	6.73e-28
27	225	27.7	862	4	Q92854	SEMAPHORIN	6.73e-28
28	224	27.6	760	11	Q62178	SEMAPHORIN B PRECURSOR	1.12e-27
29	217	26.8	284	11	Q54948	SEMAPHORIN IV HOMOLOG	3.80e-26
30	203	25.0	299	13	Q90666	COLLAPSPIN-5 (FRAGMENT)	4.09e-23
31	203	25.0	785	13	Q42337	COLLAPSPIN 5.	4.09e-23
32	168	20.7	888	11	Q35464	SEMAPHORIN VIA.	9.29e-16
33	160	19.7	771	5	Q24322	SEMAPHORIN-I.	3.94e-14
34	155	19.1	1093	11	Q60519	SEMAPHORIN G PRECURSOR	3.98e-13
35	154	19.0	886	11	Q54951	SEMAPHORIN N (SEMAPHOR	6.30e-13
36	154	19.0	887	11	O70141	SEMAPHORIN Z.	6.30e-13
37	152	18.7	494	4	Q60408	SEMAPHORIN F (FRAGMENT	1.58e-12
38	152	18.7	1074	4	Q13591	SEMAPHORIN F HOMOLOG.	1.58e-12
39	151	18.6	1077	11	Q62217	SEMAPHORIN F PRECURSOR	2.49e-12
40	149	18.4	706	5	Q24323	SEMAPHORIN-II.	6.19e-12
41	140	17.3	712	5	Q26972	SEMAPHORIN-I PRECURSOR	3.54e-10
42	135	16.6	562	5	Q17330	CESEMA.	3.22e-09
43	135	16.6	730	5	Q26473	FASCICLIN IV.	3.22e-09
44	131	16.2	770	5	O44253	SIMILAR TO SEMAPHORIN-	1.85e-08
45	110	13.6	452	5	O17816	FL5A4.1 PROTEIN.	1.21e-04

ALIGNMENTS

RESULT 1							
ID	O75326	PRELIMINARY;	PRT;	666	AA.		
AC	O75326;						
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)						
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)						
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)						
DE	SEMAPHORIN L.						
GN	SEMA.						
OS	HOMO SAPIENS (HUMAN).						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;						
OC	CATARRHINI; HOMINIDAE; HOMO.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 98389619.						
RA	LANG E C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;						
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA						
RL	viruses".						
RL	GENOMICS 51:340-350(1998).						
DR	EMBL; AF030698; G3523115; -.						
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;						

Query Match	100.0%;	Score 811;	DB 4;	Length 666;
Best Local Similarity	100.0%;	Pred. No. 6.86e-170;		
Matches	109;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Db	262	VAQLCRGDQGGESSLSVSKWTFKAMLCVSDAATNKNFNRLQDVFLPDSPSGQWRDTRV	321	
Qy	230	VAQLCRGDQGGESSLSVSKWTFKAMLCVSDAATNKNFNRLQDVFLPDSPSGQWRDTRV	289	
Db	322	YGVFSNPNYSACVYSLGDDIKVFTSSSLGKYHSSLPNPRPGKCLPDQ	370	
Qy	290	YGVFSNPNYSACVYSLGDDIKVFTSSSLGKYHSSLPNPRPGKCLPDQ	338	

RESULT 2							
ID	O88371	PRELIMINARY;	PRT;	393	AA.		
AC	O88371;						
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)						
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)						
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)						
DE	SEMAPHORIN L (FRAGMENT).						
GN	SEMA.						
OS	MUS MUSCULUS (MOUSE).						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;						

```
OC SCIURGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389619.
RA LANGE C., LJEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.:
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RL viruses.";
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030699; G3523117; -.
FT NON TER 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;

Query Match 94.8%; Score 769; DB 11; Length 393;
Best Local Similarity 94.4%; Pred. No. 2.52e-159;
Matches 101; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 260 VAQICRGDQGGESSLSVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLPPSGQWRDTRV 319
QY 230 VAQICRGDQGGESSLSVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLPPSGQWRDTRV 289
Db 320 YGVFSPNWTSAVCVYSLGIDIDRVFTSSSLKGYHMGSLNRPNGMCLP 366
QY 290 YGVFSPNWTSAVCVYSLGIDIDKVFRTSSSLKGYHSSLPNRPNGKCLP 336

RESULT 3
ID Q64906 PRELIMINARY; PRT; 653 AA.
AC Q64906;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALCELAAPHINE HERPESVIRUS 1.
OC VIRUSES; DSONA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RX MEDLINE; 97201573.
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RX MEDLINE; 97404659.
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
DR EMBL; U18243; G1000717; -.
DR EMBL; AF005370; G2337970; -.
SQ SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match 61.5%; Score 499; DB 14; Length 653;
Best Local Similarity 56.0%; Pred. No. 2.59e-92;
Matches 61; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

Db 284 VGOICRSDQGGESSLSVYKWTFLKARLACVDYDTGRIYNELODIFIWQAPENSWEETLI 343
QY 230 VAQICRGDQGGESSLSVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLPPSGQWRDTRV 289
Db 344 YGLEFLSPNWTSAVCVFVKDIDHVFETSKLKNYHHKLTPTPRPGCKMKNH 392
QY 290 YGVFSPNWTSAVCVYSLGIDIDKVFRTSSSLKGYHSSLPNRPNGKCLPDQ 338

RESULT 4
ID Q13214 PRELIMINARY; PRT; 749 AA.
```

```
AC Q13214;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V. (HUMAN).
OS HOMO SAPIENS.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96210603.
RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT cancer deletion region and demonstrate distinct expression
RT patterns.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR EMBL; U28369; G974284; -.
DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 749 AA; 83121 MW; 6138C5D5 CRC32;

Query Match 34.6%; Score 281; DB 4; Length 749;
Best Local Similarity 47.8%; Pred. No. 1.81e-40;
Matches 43; Conservative 19; Mismatches 20; Indels 8; Gaps 6;

Db 265 VGOICRDYGGQSRSL-VNKWTTFLKARLVCSVPVGEQDTHFDQLQDVFLLS-SRDH-RTP 321
QY 230 VAQICRGDQGGESSLSVSKWNTFLKAMLVCS--DAATNKNFNRLQDVFLPPSGQWRDT 287
Db 322 LLYAVFTSSSIFFQGSVAVCVYSKNDVRRAF 351
QY 288 RVYGVFSNPWN-Y--SAVCVYSLGIDIKVF 314

RESULT 5
ID Q93018 PRELIMINARY; PRT; 750 AA.
AC Q93018;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V. (HUMAN).
OS HOMO SAPIENS.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA DANTE M., WAMSLEY P.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBBJ DATA BANKS.
DR EMBL; U73167; G2880035; -.
DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 750 AA; 83034 MW; CBC87E34 CRC32;

Query Match 34.6%; Score 281; DB 4; Length 750;
Best Local Similarity 47.8%; Pred. No. 1.81e-40;
Matches 43; Conservative 19; Mismatches 20; Indels 8; Gaps 6;

Db 266 VGOICRDYGGQSRSL-VNKWTTFLKARLVCSVPVGEQDTHFDQLQDVFLLS-SRDH-RTP 322
QY 230 VAQICRGDQGGESSLSVSKWNTFLKAMLVCS--DAATNKNFNRLQDVFLPPSGQWRDT 287
Db 323 LLYAVFTSSSIFFQGSVAVCVYSKNDVRRAF 352
QY 288 RVYGVFSNPWN-Y--SAVCVYSLGIDIKVF 314
```


OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA;
OC CATARRHINI; HOMINIDAE; HOMO.
[1]
RN
RP SEQUENCE FROM N.A.

FT	CHAIN	ZI	SEMAPHORIN D.
FT	DOMAIN	593	IG-LIKE C2-TYPE DOMAIN
SQ	SEQUENCE	772 AA;	88710 MW; B929B537 CRC32;

Query Match 32.8%; Score 266; DB 11; Length 772;
Best Local Similarity 44.0%; Pred. No. 4.69e-37;
Matches 40; Conservative 24; Mismatches 18; Indels 9; Gaps 7;

```

Db      265 IGQICKNDEGGHSL-VNKWTTFLKARLICSVPGPNCIDTDFEDELQDVLM-NSKDP-KN 321
        ::::: | || |::||| |::||| : : : |:| |::||| : :
QY      230 VAQLCRGDQGGSLSVSKWNTFLKAMLVCS--DA-ATKNKFNRLOQVFLLPDPSGQWRD 286

```

```

Db      322 PIVYGVFTTSSNIFKGSVCMYSMSDVRVF 352
        |||||: : : |||||: : : ||
Qy      287 TRVYGVFSNPWN-Y--SVCVYISLGDIDKVF 314

```

RESULT 13
ID Q63548
PRELIMINARY:
PRT: 772 AA.

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN III/COLLAPSPIN-1.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR; TISSUE-BRAIN;
RL GIGER R. I.;
RA SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X95286; E220165; -;
DR PFAM; PF00047; ig; 1.
SQ SEQUENCE 772 AA; 88808 MW; 7E88FD35 CRC32;

Query Match 32.8%; Score 266; DB 11; Length 772;
Best Local Similarity 44.0%; Pred. NO. 4.69e-37;
Matches 40; Conservative 24; Mismatches 18; Indels

```
Db      265 IGOICKNDFFGHRSL-VNKWTTFLKARLICSVPGPNGIDTHFDELQDVFLM-NSKDP-KN 321
        ::::: | | | | | | | | | | : : : | | | | | : : :
Qy      230 VAQLCRDGGOGGESSLSVKWNTFLKAMLVCS--DA-ATKNKFNRLOQVFLLPDSFGWRD 286
```

```

Db      322 PIVYGVETSSNIFKGSVAVCMYMSDVRVF 352
          ||||| : | : ||||| : : : ||
Qy      287 TRVYGVESNPWN-Y--SAVCVYSLGIDIKVF 314

```

RESULT	14
ID	Q90607
AC	Q90607;
DT	01-NOV-1996 (TREMBUREL, 01, CREATED)
DT	01-NOV-1996 (TREMBUREL, 01, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBUREL, 08, LAST ANNOTATION UPDATE)
DE	COLLAPSN.
DI	
OS	GALLUS GALLUS (CHICKEN).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES
OC	NEOGNATHIA; GALLIFORMES; PHASIANIDAE; PHASININAE; GALLUS.
RN	[1]
RP	SEQUENCE FROM N.A.

RA LUO Y., RABLE D., RAPER J.A.;
 RT "Collapsin: a protein in brain that induces the collapse and
 paralysis of neuronal growth cones."; *Cell* 75:217-227(1993).
 RL CELL 75:217-227(1993).
 DR EMBL: U02528; G410079; -.
 SF PFAM: PF00047; ig: 1.
 SQ SEQUENCE 772 AA; 88867 MW; BC6CB9E8 CRC32:

Query Match 32.6%; Score 264; DB 13; Length 772;
Best Local Similarity 44.0%; Pred. No. 1.33e-36;
Matches 40; Conservative 23; Mismatches 19; Indels

Db 265 IGQICKNDFGHRSL-VNKWTTFLKARLICSVPGPNGIDTHFDELDQDVFLM-NSKDP-KN 321

```
QY      230 VAQLRGDQGGESSLVS/KWNTFLKAWLVC--DA-ATKNFNRLQDVFLPDPGQWRD 286  
Db      322 PIYGVFTTSSNIFKGSACVMSWTDDRRVF 352  
QY      287 TRYVGYSFPWN-Y--SACVYSLGDIDKFV 314
```

RESULT 15
ID 062179
PRELIMINARY:
PRT: 782 AA

062179;
AC 01-NOV-1998 (TREMBREL. 08, CREATED)
DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN C (SEM C) (FRAGMENT).
GN SEMAC OR SEMC.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI; TISSUE-BRAIN;
RX MEDLINE; 95267431.
RA PUESCHEL A. W., ADAMS R. H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC LOW LEVELS FOUND BETWEEN DAYS 10-12.
CC EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL
CC BIRTH.

CC	-!	SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC	--	CONTAINS ONE C2-LIKE DOMAIN.
DR	EMBL:	X85992; G854328;
DR	MGD:	MGI:107559; SEMAC.
KW	IMMUNOGLOBULIN FOLD;	MULTIGENE FAMILY; NEUROGENESIS;
DR	IMMUNOGLOBULIN FOLD;	MULTIGENE FAMILY; NEUROGENESIS;
KW	DEVELOPMENTAL PROTEIN.	
KW	NON_TER	1
FT	DOMAIN	550 607 IG-LIKE C2-TYPE DOMAIN.
SC	SEQUENCE	782 AA; 86823 MW; CC80E95D CRC32:
SQ		

Query Match	31.8%;	Score 258;	DB 11;	Length 782;
Best Local Similarity	45.1%;	Pred. No. 3.01e-35;		
Matches	41;	Conservative	19;	Mismatches 23;
				Indels: 8;
				Gaps: 5;

```

228 VARVCKGDEGGERVLQ-QRWTSFLKAQLLCSRPDGDPFNVLQDVETLNPQ-DWRKTL 285
||:|:|:|:| | :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
230 VAQLCRGDGGESLSVSKWNTFLKAMLVCSDAATNKNENRLODVFL-PDPGGWDRTR 288

```

Db 286 SIGVFTSQWHRGTTTEGSAICVFTMNDVQKAF 316
 ||||: |: ||:||||: |:|:|
 QY 289 VYGVFSPWN--Y--SACVYSLGDIDKVF 314

Search completed: Thu Jul 8 18:38:30 1999
Job time : 15 secs.

THIS PAGE BLANK (USPTO)

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:41:26 1999; Maspar time 10.57 Seconds

Tabular output not generated. 18.111 Million cell updates/sec

Title: >US-09-041-236-2
Description: (237-245) from US09041236.pep (17 of 45)

Perfect Score: 53
Sequence: 1 DQGGSSLS 9

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 14.296; Variance 36.534; scale 0.391

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	48	90.6	966	34	Solanum tuberosum tub	1.80e+01
2	42	79.2	245	16	Hepatitis GB virus (H	1.06e+02
3	41	77.4	441	13	Vaccinia virus semaph	1.41e+02
4	40	75.5	452	5	EPSP synthase wild ty	1.88e+02
5	39	73.6	382	6	Mannitol-1 dehydrogen	2.50e+02
6	39	73.6	516	3	Soybean glycinin A3B4	2.50e+02
7	39	73.6	1411	20	Nucleolar/endosomal a	2.50e+02
8	38	71.7	319	28	S. marcescens Ssp auto	3.31e+02
9	38	71.7	319	28	S. marcescens Ssp auto	3.31e+02
10	38	71.7	355	25	Human clone 56 protei	3.31e+02
11	38	71.7	378	29	Aplysia cAMP-response	3.31e+02
12	38	71.7	410	32	Decaprenyl diphosphat	3.31e+02
13	38	71.7	485	20	Arabidopsis S-adenosyl	3.31e+02
14	38	71.7	485	20	Asparagus S-adenosyl-	3.31e+02
15	38	71.7	486	5	Prod. of Nicotiana ta	3.31e+02
16	37	69.8	43	6	pH-dependent membrane	4.37e+02

17	37	69.8	44	1	R03437	pH-dependent membrane	4.37e+02
18	37	69.8	320	39	W89744	Staphylococcus aureus	4.37e+02
19	37	69.8	432	23	W14081	S.thermophilus exopol	4.37e+02
20	37	69.8	473	23	W22180	S.thermophilus exopol	4.37e+02
21	37	69.8	682	17	R87154	Alternatively spliced	4.37e+02
22	37	69.8	682	11	R49144	Product of alternativ	4.37e+02
23	37	69.8	836	11	R58912	Product of alternativ	4.37e+02
24	37	69.8	836	17	R87153	Alternatively spliced	4.37e+02
25	37	69.8	904	17	R87147	Protocadherin clone 4	4.37e+02
26	37	69.8	904	11	R58907	Human protocadherin-4	4.37e+02
27	37	69.8	1462	7	R37508	Human DNA polymerase	4.37e+02
28	37	69.8	1470	28	W23411	Porcine transmissible	4.37e+02
29	37	69.8	1805	11	R60126	Rat nestin protein is	4.37e+02
30	37	69.8	1805	5	R27204	Rat nestin.	4.37e+02
31	36	67.9	340	37	W81594	Protein encoded by hu	5.76e+02
32	36	67.9	398	24	W24095	L-methionine gamma-Ly	5.76e+02
33	36	67.9	398	15	R80470	Pseudomonas putida L-	5.76e+02
34	36	67.9	398	20	W10236	Pseudomonas putida me	5.76e+02
35	36	67.9	520	22	W15747	Progestin-regulated g	5.76e+02
36	36	67.9	550	8	R42254	Human p65 protein DNA	5.76e+02
37	36	67.9	550	8	R42085	NF-KappaB p65 protein	5.76e+02
38	36	67.9	550	8	R42255	Human p65 protein DNA	5.76e+02
39	36	67.9	686	4	P93711	Sequence of pseudorab	5.76e+02
40	36	67.9	797	38	W85043	NFKappaB p65 subunit-	5.76e+02
41	36	67.9	797	38	W85042	PKB-green fluorescent	5.76e+02
42	36	67.9	1055	23	W13055	HIV-2 provirus-encode	5.76e+02
43	36	67.9	1717	27	W23331	Neuroblastoma indicat	5.76e+02
44	36	67.9	1831	27	W23329	Microtubule-associate	5.76e+02
45	36	67.9	2670	15	R88125	Rat IP3 receptor.	5.76e+02

ALIGNMENTS

RESULT 1
ID W49077 standard; Protein; 966 AA.
AC W49077;
DT 09-NOV-1998 (first entry)
DE Solanum tuberosum tuber L-type alpha glucan phosphorylase.
KW L-type; tuber; alpha-glucan phosphorylase; cold storage; chips;
KW cold-sweetening.
OS Solanum tuberosum.
FH Key Location/Qualifiers
FT Peptide 1..50 /note= "signal peptide"
FN W09835051-A1.
PD 13-AUG-1998.
PF 05-FEB-1998; CA0055.
PR 04-JUN-1997; US-868786.
PR 10-FEB-1997; US-036946.
PA (MIAC) CANADA DEPT AGRIC & AGRI-FOOD CANADA.
PI Armstrong JD, Kawchuk LM, Knowles NR, Lynch DR;
DR WPI; 98-447248/38.
DR N-PSDB; V32918.
PT Transgenic potato plants with improved cold storage characteristics
PT - are modified plants with reduced alpha glucan L or H type
PT phosphorylase activity, and have reduced cold sweetening
PS Disclosure; Page 50-53; 9ipp; English.
CC The sequence is that of alpha-glucan L-type tuber phosphorylase
CC which can be used in the production of a potato plant with
CC improved tuber cold-storage characteristics. Such plants can be
CC used in agricultural potato production. The potatoes are especially
CC useful in the manufacture of potato chips. The method reduces
CC cold-sweetening (conversion of starches to sugars), resulting
CC in the ability to store potatoes cooler temperatures. This
CC produces prolonged dormancy, reduced incidence of disease and
CC increased storage life. The potatoes also have improved specific
CC gravity and improved chip score.
SQ Sequence 966 AA;

Query Match 90.6%; Score 48; DB 34; Length 966;
Best Local Similarity 77.8%; Pred No. 1.80e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Db 62 egggsdls 70
Qy 237 DQGGESSLS 245

RESULT 2
ID R81413 standard; Protein; 245 AA.
AC R81413;
DE Hepatitis GB virus (HGBV) clone 2 protein prod.
KW Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;
KW reagents; non-A; non-B; non-C; non-D; non-E; clone 2;
OS tamatin; infected plasma; lambda phage; cDNA library.
FH Hepatitis GB virus.
FT Key Location/Qualifiers
FT misc_difference 59
FT misc_difference 91 /note= "corresponding codon STOP codon"
FT misc_difference 91 /note= "corresponding codon STOP codon"
FN WO9521922-A2.
PD 17-AUG-1995.
PF 14-FEB-1995; U02118.
PR 14-FEB-1994; US-196030.
PR 13-MAY-1994; US-242654.
PR 29-JUL-1994; US-283314.
PR 23-NOV-1994; US-344190.
PR 23-NOV-1994; US-344185.
PR 27-JAN-1995; US-344557.
PA (ABBO ) ABBOTT LAB
PI Buljk SL, Dawson GJ, Desai SM, Erker JC, Leary TP;
PI Muerhoff AS, Mushahwar IK, Pilot-Matias TJ, Schlauder GG;
PI Simons JN;
DR WPI: 95-293123/38.
DR N-PSDB: T00042.
PT Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful
PT for diagnosis and therapy of hepatitis GB virus
PS Example 5; Pages 199-200; 661pp; English.
CC Double stranded hepatitis GB virus (HGBV) DNA obtd. from HGBV
CC infected tamarin plasma, using standard procedures, was used to
CC prepare a lambda phage HGBV cDNA library. The cDNA clone T00042,
CC which encodes the proteins R81411-16 (the 6 possible reading a
CC frames), was rescued from the lambda phage, searched against a
CC sequence database and found to be an unique HGBV sequence.
CC Reagents which comprise the HGBV DNA, or its protein prods. can
CC be used for the diagnosis, therapy or in a vaccine to prevent
CC HGBV infection.
SQ Sequence 245 AA;

Query Match 79.2%; Score 42; DB 16; Length 245;
Best Local Similarity 87.5%; Pred. NO. 1.06e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 197 qggssls 204
Qy 238 QGGESSLS 245

RESULT 3
ID R71381 standard; Protein; 441 AA.
AC R71381;
DE Vaccinia virus semaphorin IV protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Vaccinia virus.
FN WO9507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC ) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;

PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB: Q87443.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 65-67; 101pp; English.
CC The sequence of the vaccinia virus semaphorin IV protein.
CC The gene sequence was isolated as the A39R open reading frame sequence
CC from variola, based on sequence homology searches of a database with the
CC grasshopper, Tribolium and Drosophila semaphorin sequences. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III
CC (Q87442), vaccinia virus semaphorin IV, Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 441 AA;

Query Match 77.4%; Score 41; DB 13; Length 441;
Best Local Similarity 77.8%; Pred. NO. 1.41e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 236 deggssls 244
Qy 237 DQGGESSLS 245

RESULT 4
ID R27797 standard; Protein; 452 AA.
AC R27797;
DE EPSP synthase (first entry)
DE EPSP synthase wild type from Aspergillus.
KW 5-enolpyruvyl-3-phosphoshikimate synthase; Glyphosate; tolerant;
KW mutant; weeds; herbicide.
OS Aspergillus.
PN US5145783-A.
PD 08-SEP-1992.
PF 26-MAY-1987; 054337.
PR 26-MAY-1987; US-054337.
PR 22-APR-1988; US-179245.
PR 09-JUL-1990; US-550276.
PA (MONS ) MONSANTO CO.
PI Kishore GM, Shah DM;
PI WPI: 92-339628/41.
DR Glyphosate-tolerant plants and seeds - contg. DNA encoding
DR mutant glyphosphate tolerant 5-enol:pyruvyl-3-phospho:shikimate
PT (EPSP) synthase
PS Disclosure; Fig 2; 34pp; English.
CC EPSP synthase from Aspergillus was used in a comparison of
CC EPSP synthase enzymes from plant, bacterial and fungal species to
CC determine conserved regions. The gene encoding EPSP synthase may be
CC altered by oligonucleotide mutagenesis to insert the DNA substn.
CC necessary to result in an alanine for glycine substn. in the conserved
CC amino acid sequence L-G-N-A-G-T-A. Introduction of the alanine for
CC glycine substn. makes EPSP synthase glyphosate tolerant.
CC See also R27794-804.
SQ Sequence 452 AA;

Query Match 75.5%; Score 40; DB 5; Length 452;
Best Local Similarity 55.6%; Pred. NO. 1.89e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 78 ehgggstls 86
Qy 237 DQGGESSLS 245

RESULT 5
ID R28826 standard; Protein; 382 AA.

```



```

AC R28826;
DT 26-MAR-1993 (first entry)
DE Mannitol-1 dehydrogenase.
KW M1PD; polyol; biosynthesis; plant; Intl;
KW myo-inositol O-methyl transferase.
OS Escherichia coli mtd mutant 239.
PN W09219731-A.
PD 12-NOV-1992.
PF 08-MAY-1992; U03826.
PR 09-MAY-1991; US-697390.
PR 20-APR-1992; US-871416.
PA (UVAR-) UNIV ARIZONA.
PI Bohner HJ, Jensen RG, Tarczynski MC, Vernon DM;
DR WPI; 92-398058/48.
DR N-PSDB; Q31199.
PT Transgenic plants having improved stress tolerance and strength -
PT contain foreign gene conditioning for expression of enzyme
PT catalysing prodn. of polyol in plant tissue from sugars
PS Claim 22: Page 47 + 36-37; 55pp; English.
CC Two separate genes are capable of inducing novel polyol biosynthesis
CC in transgenic plants. One gene is bacterial in origin, e.g. E. coli
CC derived mtd encoding mannitol-1-P dehydrogenase (Q31199). The gene
CC for this enzyme has been described by Lee and Sailer, J. Bact., 153:2,
CC 685-692 (1983). The other gene is ImI encoding myo-inositol O-
CC methyl transferase, derived from a stress tolerant plant (Q31200).
SQ Sequence 382 AA;

Query Match 73.6%; Score 39; DB 6; Length 382;
Best Local Similarity 55.6%; Pred. No. 2.50e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 108 eqgnesp1n 116
QY 237 DQGESLS 245

RESULT 6
ID P61362 standard; Protein; 516 AA.
AC P61362;
DT 16-OCT-1991 (first entry)
DE Soybean glycinin A3B4 subunit.
KW Soybean protein; glycinin.
OS Glycine max.
PN J61132189-A.
PD 19-JUN-1986.
PF 03-DEC-1984; 254217.
PR 03-DEC-1984; JP-254217.
PA (NORO) NORIINSHO KK.
DR WPI; 86-200345/31.
DR N-PSDB; N60939.
PT Prepn. of soybean messenger RNA - for insertion into cells or
PT microorganisms to produce soybean protein.
PS Example 1; Fig 1; 7pp; Japanese.
CC Sequence derived from mRNA may be used for the expression of the
CC soybean protein by a foreign host.
SQ Sequence 516 AA;

Query Match 73.6%; Score 39; DB 3; Length 516;
Best Local Similarity 62.5%; Pred. No. 2.50e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 455 egggeqgl 462
QY 237 DQGESLS 244

RESULT 7
ID W02258 standard; Protein; 1411 AA.
AC W02258;
DT 09-MAR-1997 (first entry)
DE Nucleolar/endosomal auto-antigen p162.
KW Auto-antibody; p162; rheumatic disease; antigen; diagnosis;
KW gene therapy.

```

```

OS Homo sapiens.
PN DE1951514-CI.
PD 12-SEP-1996.
PR 27-APR-1995; 015514.
PR 27-APR-1995; DE-015514.
PA (PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.
PI Renz M, Seelig HP;
DR WPI; 96-403153/41.
DR N-PSDB; T58751.
PT DNA encoding nucleolar-endosomal auto-antigen - useful for exact
PT diagnosis of rheumatic disease, in gene therapy and for removal of
PT specific auto-antibodies
PS Claim 1; Fig 2; 15pp; German.
CC Transformed cells can be cultured to produce the antigen p162, for use
CC in exact (differential) diagnosis of rheumatic disease, i.e. they
CC can detect, in immunoassays, Western blots, etc., rheumatism-
CC specific auto-antibodies. The antigen can be used therapeutically,
CC in the removal of auto-antibodies from the circulation, or when
CC coupled to a cytotoxin, the elimination of auto-antibody-
CC producing lymphocytes.
SQ Sequence 1411 AA;

Query Match 73.6%; Score 39; DB 20; Length 1411;
Best Local Similarity 62.5%; Pred. No. 2.50e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 72 hggesnla 79
QY 238 QGGESLS 245

RESULT 8
ID W27716 standard; Protein; 319 AA.
AC W27716;
DT 08-MAY-1998 (first entry)
DE S. marcescens Ssp autotransporter region.
KW Ssp protein; autotransporter; diagnostic; therapy;
KW Gram-negative bacteria; surface presented polypeptide.
OS Serratia marcescens.
FH Key Location/Qualifiers
FT Protein 1..319 /note= "partial protein sequence"
FN WO9735022-AI.
PD 25-SEP-1997.
PR 15-MAR-1996; E01130.
PR 15-MAR-1996; WO-E01130.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Jose J, Maurer J, Meyer TF;
DR WPI; 97-480227/44.
DR N-PSDB; T88153.
PT Presentation of peptide(s) on surface of Gram-negative bacteria -
PT via transformation with vector encoding signal peptide, presented
PT peptide and transporter domain of auto-transporter, producing
PT peptide libraries for epitope mapping
PS Claim 7; Fig 20; 84pp; German.
CC This sequence represents an autotransporter membrane integration region
CC from the S. marcescens IFO-3046 Ssp gene. This region is involved in a
CC novel method which allows the presentation of stable fusion polypeptides
CC on the surface of Gram-negative bacteria which can be released into the
CC surrounding media. The method can be used to produce a variegated
CC population of surface-presented polypeptides, so that bacteria expressing
CC polypeptides with particular properties can be identified and
CC simultaneously selected, e.g. for epitope mapping or selection of ligands
CC with the highest affinity for antibodies, major histocompatibility
CC complex (MHC) molecules or other components of the immune system.
CC Selected polypeptides can be used diagnostically, e.g. to screen sera or
CC antibody banks, and polypeptide expressing cells may be used as live
CC vaccines. They may be used therapeutically, e.g. when the polypeptide is
CC an antibody, to remove or concentrate pollutants, inactivate toxins,
CC prepare and process food, prepare washing compositions and label cells.
CC Selected bacteria can be stored, reproduced and replicated on a large
CC scale as individual clones.
SQ Sequence 319 AA;

```

Query Match 71.7%; Score 38; DB 28; Length 319;
Best Local Similarity 50.0%; Pred. No. 3.31e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 200 egggaal 207
QY 237 DQGGESSL 244

RESULT 9
ID W27715 standard; Protein; 319 AA.
AC W27715;
DT 08-MAY-1998 (first entry)
DE S. marcescens Ssp autotransporter membrane integration region.
KW Ssp protein; autotransporter; diagnostic; therapy;
OS Gram-negative bacteria; surface presented polypeptide.
FH Key Location/Qualifiers
FT Protein 1..319
FT W09735022-A1. /note= "partial protein sequence"
PD 25-SEP-1997.
PF 15-MAR-1996; E01130.
PR 15-MAR-1996; WO-E01130.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Jose J, Maurer J, Meyer TF;
DR WPI; 97-480227/44.
DR N-PSDB; T88152.
PT Presentation of peptide(s) on surface of Gram-negative bacteria -
via transformation with vector encoding signal peptide, presented
PT peptide and transporter domain of auto-transporter, producing
PT peptide libraries for epitope mapping
PS Claim 7; Fig 19; 84pp; German.
CC This sequence represents an autotransporter membrane integration region
from the S. marcescens RH1 Ssp gene. This region is involved in a novel
method which allows the presentation of stable fusion polypeptides on
the surface of Gram-negative bacteria which can be released into the
surrounding media. The method can be used to produce a variegated
population of surface-presented polypeptides, so that bacteria expressing
polypeptides with particular properties can be identified and
simultaneously selected, e.g. for epitope mapping or selection of ligands
with the highest affinity for antibodies, major histocompatibility
complex (MHC) molecules or other components of the immune system.
CC Selected polypeptides can be used diagnostically, e.g. to screen sera or
antibody banks, and polypeptide expressing cells may be used as live
vaccines. They may be used therapeutically, e.g. when the polypeptide is
an antibody, to remove or concentrate pollutants, inactivate toxins,
prepare and process food, prepare washing compositions and label cells.
CC Selected bacteria can be stored, reproduced and replicated on a large
scale as individual clones.
SQ Sequence 319 AA;

Query Match 71.7%; Score 38; DB 28; Length 319;
Best Local Similarity 50.0%; Pred. No. 3.31e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 200 egggaal 207
QY 237 DQGGESSL 244

RESULT 10
ID W05398 standard; Protein; 355 AA.
AC W05398;
DT 19-FEB-1998 (first entry)
DE Human clone 56 protein.
KW Src-homology region 3 domain; mouse; SH3 domain; cell growth;
cellular signalling element; cellular structural element; malignancy;
protein identification; functional domain; protein screening;
cellular signal transduction process.
OS Homo sapiens.
PN W09631625-A1.

PD 10-OCT-1996.
PF 04-APR-1996; U04454.
PR 03-APR-1996; US-630915.
PR 07-APR-1995; US-4117872.
PA (CYTO-) CYTOGEN CORP.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Hoffman N, Ray BK, Mcconnell SJ, Sparks AB;
DR WPI; 96-465045/46.
DR N-PSDB; T39798.
PT Identifying polypeptide(s) having specific functional domain (esp.
SH3 domain) - comprises detecting selective binding to recognition
unit, regardless of sequence homology
PS Claim 54; Fig 57; 174pp; English.
CC W0386-W03403 represent novel human and mouse Src-homology region 3 (SH3)
domain containing proteins that can be used in the method of the
invention. SH3 domain containing proteins play a role in signalling and
structural elements of cells. The method of the invention is for
identifying polypeptides containing functional domains of interest
(especially SH3 domains). The method comprises contacting a multivalent
recognition unit (RU) complex with a number of peptides and identifying
polypeptides having a selective binding affinity for the RU complex. The
method is based on functional similarities and does not rely on sequence
similarities. Prior methods only gave limited success for identifying
proteins which contain an SH3 domain due to the minimal sequence homology
among known SH3 proteins. It has been found that small peptide RUS in
multivalent form have reduced specificity for a given functional domain
compared to monomer RUS. Multivalent RU complexes are particularly suited
to screening for polypeptides containing functional domains that are
similar to, but not identical in sequence to, the original target
functional domain. The new method enables proteins having a common
function to be identified. Identification of novel SH3 proteins will be
useful for a better understanding of cell growth, malignancy, signal
transduction processes, etc. New candidate drugs can be identified, and
their specificities (e.g. pharmacological activities) can be assessed
using the method of the invention.
SQ Sequence 355 AA;

Query Match 71.7%; Score 38; DB 25; Length 355;
Best Local Similarity 71.4%; Pred. No. 3.31e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 89 egggeas 95
QY 237 DQGGESS 243

RESULT 11
ID W41508 standard; Protein; 378 AA.
AC W41508;
DT 22-JUN-1998 (first entry)
DE Aplysia cAMP-response element binding protein 2.
KW APCR2B-2; cAMP-response element binding protein-2; snail;
memory loss; Alzheimer's disease; amnesia; ischaemia; head trauma;
neuronal injury; Parkinson's disease; senility; therapy.
OS Aplysia californica.
FH Key Location/Qualifiers
FT Region 73..108
FT /note= "leucine heptad repeat region"
FT Modified_site 150..153
FT /note= "MAP kinase phosphorylation site"
FT Modified_site 234..237
FT /note= "MAP kinase phosphorylation site"
FT Modified_site 271..174
FT /label= "protein kinase C phosphorylation site"
FT Domain 306..378
FT /note= "basic region/leucine zipper domain,
interacts with ApC/EBP"

PN W09746257-A1.
PD 11-DEC-1997.
PF 03-JUN-1997; U09438.
PR 03-JUN-1996; US-656811.
PA (UYCO) UNIV COLUMBIA NEW YORK.
PI Bartsch D, Ghirardi M, Kandel ER;

DR WPI: 98-051903/05.
PT N-PSDB: V04079.
PT Enhancing long-term memory in subjects whose cAMP-responsive gene is
PT repressed - used to treat long-term memory defects, e.g. age-related
PT memory loss, Alzheimer's disease
PS Example 2; Page 73-74; 100pp; English.
CC This protein comprises the cAMP-response element binding protein-2,
CC i.e. ApCREB-2, of the marine snail *Aplysia*. ApCREB-2 is a
CC transcription factor and repressor of long-term facilitation in
CC *Aplysia* neurons. It is constitutively expressed in sensory
CC neurons. ApCREB-2 is a homologue of human CREB-2 and mouse ATF-4.
CC The amino acid sequence of ApCREB-2 was deduced from a polynucleotide
CC sequence (see V04079) obtained from 2 clones isolated from an *Aplysia*
CC central nervous system cDNA library. The invention provides a method
CC of enhancing long-term memory in a subject whose cAMP-responsive gene
CC expression is repressed due to binding of a CREB-2 to a protein or
CC DNA associated with cAMP-responsive gene expression, or both. The
CC method involves administering to the subject a compound capable of
CC interfering with such binding so as to derepress cAMP-responsive gene
CC expression and thereby enhance long-term memory. Such compounds
CC include anti-CREB-2 antibodies or a compound capable of altering
CC phosphorylation of CREB-2. The method is used to treat e.g.
CC age-related memory loss, Alzheimer's disease, amnesia, ischaemia,
CC shock, head trauma, neuronal injury, toxicity or degradation,
CC Parkinson's disease or senility (claimed).
SQ Sequence 378 AA;

Query Match 71.7%; Score 38; DB 29; Length 378;
Best Local Similarity 62.5%; Pred. No. 3.31e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 38 hggdesls 45
|||:|
QY 238 QGESSLS 245

RESULT 12
ID W53921 standard; Protein; 410 AA.
AC W53921;
DT 19-AUG-1998 (first entry)
DE Decaprenyl diphosphate synthase #2.
KW Decaprenyl diphosphate synthase; DBS; ubiquinone-10; transgenic;
KW microbial; host.
OS Gluconobacter suboxydans.
FH Key Location/Qualifiers
FT Misc_difference 339
FT /note= "stop codon"
FT Misc_difference 366
FT /note= "stop codon"
FT Misc_difference 372
FT /note= "stop codon"
FT J10057072-A.
PN 03-MAR-1998.
PD 03-MAR-1998.
PF 22-AUG-1996; 238682.
PR 22-AUG-1996; JP-238682.
PA (ALPH-) ALPHA SHOKUHIN KK.
DR WPI: 98-210404/19.
DR N-PSDB; V23681.
PT Production of highly pure ubiquinone-10 - by recombinant
PT microorganism comprising gene encoding decaprenyl di-phosphate
PT synthase
PS Claim 5; Page 10-11; 14pp; Japanese.
CC The present sequence represents decaprenyl diphosphate synthase (DBS)
CC from *Gluconobacter suboxydans*. The protein is shown to continue after
CC the first stop codon but the claimed sequence having 315 amino acids
CC does not. Extended protein is shown in W53920 and W53921, while the
CC 315 amino acid sequence as shown in figure 1 of the specification is
CC shown in W53922. The present invention describes a transgenic microbial
CC organism comprising a gene encoding DBS. The transgenic microbial host
CC may be used to produce ubiquinone-10. The ubiquinone-10 produced by the
CC host is highly pure.
SQ Sequence 410 AA;

Query Match 71.7%; Score 38; DB 32; Length 410;
Best Local Similarity 85.7%; Pred. No. 3.31e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 8 ggesals 14
|||||
QY 239 GGESLS 245

RESULT 13
ID W01459 standard; Protein; 485 AA.
AC W01459;
DT 23-FEB-1997 (first entry)
DE Arabidopsis S-adenosyl-L-homocysteine hydrolase.
KW S-adenosyl-L-homocysteine hydrolase; SHH; promoter; monocot; dicot;
KW transgenic plant; disease resistance; pathogen resistance.
OS Arabidopsis thaliana.
PN W09632488-A1.
PD 17-OCT-1996.
PF 10-APR-1996; G00882.
PR 10-APR-1995; GB-007381.
PA (ZENE) ZENECA LTD.
PI Draper J, Greenland AJ, Skipsey M, Warner S;
DR WPI: 96-477138/47.
DR N-PSDB: T44515.
PT S-adenosyl-L-homocysteine hydrolase promoter - used for driving
PT expression of effector genes, such as pathogen resistance genes, in
PT transgenic plants
PS Example 4; Fig 6; 57pp; English.
CC Arabidopsis S-adenosyl-L-homocysteine hydrolase (SHH) (W01459)
CC catalyses the reversible hydrolysis of S-adenosyl-L-homocysteine
CC to adenosine and homocysteine. Its amino acid sequence was
CC deduced from the SHH gene (T44515) isolated from an Arabidopsis
CC genomic DNA library. The promoter of the Arabidopsis SHH gene is
CC useful for the expression of effector genes in transgenic plants.
SQ Sequence 485 AA;

Query Match 71.7%; Score 38; DB 20; Length 485;
Best Local Similarity 50.0%; Pred. No. 3.31e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 138 dggdatl 145
|.:|.:|
QY 237 DQGESL 244

RESULT 14
ID W01458 standard; Protein; 485 AA.
AC W01458;
DT 22-FEB-1997. (first entry)
DE Asparagus S-adenosyl-L-homocysteine hydrolase.
KW S-adenosyl-L-homocysteine hydrolase; SHH; promoter; monocot; dicot;
KW transgenic plant; disease resistance; pathogen resistance.
OS Asparagus officinalis.
FH Key Location/Qualifiers
FT region 150..190
FT /note= "region found in SHH of photosynthetic
FT species"
FT binding_site 263..294
FT /label= NAD+-binding_site
FT W09632488-A1.
PN 17-OCT-1996.
PD 17-OCT-1996.
PF 10-APR-1996; G00882.
PR 10-APR-1995; GB-007381.
PA (ZENE) ZENECA LTD.
PI Draper J, Greenland AJ, Skipsey M, Warner S;
DR WPI: 96-477138/47.
DR N-PSDB: T44513.
PT S-adenosyl-L-homocysteine hydrolase promoter - used for driving
PT expression of effector genes, such as pathogen resistance genes, in
PT transgenic plants
PS Example 1; Fig 1; 57pp; English.
CC Asparagus S-adenosyl-L-homocysteine hydrolase (SHH) (W01458)

CC catalyses the reversible hydrolysis of S-adenosyl-L-homocysteine
CC to adenosine and homocysteine. It contains an extra stretch of
CC amino acid residues (positions 150-190) previously found in other
CC photosynthetic species, parsley and Rhodospirillum rubrum, but not
CC in SHHs from non-photosynthetic species. A cDNA sequence (T44513)
CC coding for the asparagus was used to identify the Arabidopsis
CC thaliana SHH gene (T44515) and promoter (T44514), useful for
CC expression of effector genes in transgenic plants.
SQ Sequence 485 AA;

Query Match 71.7%; Score 38; DB 20; Length 485;
Best Local Similarity 50.0%; Pred. No. 3.31e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 138 dgggdttl 145
|:|:|:|:
QY 237 DGGESSL 244

RESULT 15
ID R26500 standard; Protein; 486 AA.
AC R26500:
DT 10-MAR-1993 (first entry)
DE Prod. of Nicotiana tabacum gene expressing at floral differentiation.
KW Flower; induction.
OS Nicotiana tabacum.
PN J04258292-A.
PD 14-SEP-1992.
PF 14-FEB-1991; 020702.
PR 14-FEB-1991; JP-020702.
PA (NIBS) JAPAN TOBACCO INC.
DR WPI: 92-354683/43.
DR N-PSDB: 029419.
PT Gene expressing at floral differentiation for flowering control -
PT obtd. from culture of floral axis epithelium cells of Nicotiana
PT tabacum, for introduction into other plants
PS Disclosure; Page 6; 8pp; Japanese.
CC The protein sequence was deduced from the DNA sequence of a clone
CC obtd. by screening a cultured epithelial axis cDNA library with probes
CC from cultured and untreated mRNA. The clone obtd. expressed its
CC protein only at floral differentiation. The gene can be introduced
CC into other plants or can be suppressed by an antisense technique for
CC the control of flowering of plants.
SQ Sequence 486 AA;

Query Match 71.7%; Score 38; DB 5; Length 486;
Best Local Similarity 50.0%; Pred. No. 3.31e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 139 dgggdatl 146
|:|:|:|:
QY 237 DGGESSL 244

Search completed: Thu Jul 8 18:41:45 1999
Job time : 19 secs.

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:40:30 1999; MasPar time 3.24 Seconds
Tabular output not generated. 78.642 Million cell updates/sec

Title: >US-09-041-236-2
Description: (237-245) from US09041236.pep (17 of 45)
Perfect Score: 53
Sequence: 1 DOGGESSL5 9

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 20.517; Variance 18.179; scale 1.129

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	48	90.6	966	1	PHS1_SOLTU	5.45e-02
2	41	77.4	93	1	YGGJ_ERWCH	5.01e-00
3	41	77.4	249	1	CREB_CHLVR	5.01e-00
4	41	77.4	346	1	ASPG_MOUSE	5.01e-00
5	41	77.4	403	1	VA39_VACCC	5.01e-00
6	41	77.4	441	1	VA39_VACCV	5.01e-00
7	41	77.4	534	1	VLI_HPV41	5.01e-00
8	41	77.4	796	1	YH04_YEAST	5.01e-00
9	41	77.4	835	1	FASD_ECOLI	5.01e-00
10	41	77.4	1120	1	ROM_SCHPO	5.01e-00
11	41	77.4	2137	1	SCB_HUMAN	5.01e-00
12	41	77.4	2411	1	DAB_DROME	5.01e-00
13	40	75.5	274	1	RRPP_VSVJM	9.14e-00
14	40	75.5	274	1	RPP_VSVJM	9.14e-00
15	40	75.5	385	1	YDEM_ECOLI	9.14e-00
16	40	75.5	653	1	NUAM_DROME	9.14e-00
17	40	75.5	672	1	NQ03_PARDE	9.14e-00
18	40	75.5	671	1	NUAM_ACACA	9.14e-00
19	40	75.5	695	1	NUAM_RECAM	9.14e-00
20	40	75.5	727	1	NUAM_BOVIN	9.14e-00
21	40	75.5	727	1	NUAM_HUMAN	9.14e-00
22	40	75.5	738	1	NUAM_SOLTU	9.14e-00
23	40	75.5	744	1	NUAM_NEUCR	9.14e-00

24 39 73.6 224 1 YOCX_MYCTU HYPOTHETICAL PROTEIN C 1.65e+01
25 39 73.6 336 1 XYLZ_PSEPU TOLUATE 1,2-DIOXIGENAS 1.65e+01
26 39 73.6 382 1 MTLD_ECOLI MANNITOL-1-PHOSPHATE S 1.65e+01
27 39 73.6 434 1 PDP_BACSU PYRIMIDINE-NUCLEOSIDE 1.65e+01
28 39 73.6 452 1 F26_YEAST FRUCTOSE-2,6-BISPHOSPH 1.65e+01
29 39 73.6 516 1 GLC5_SOYBN GLYCININ PRECURSOR [CO 1.65e+01
30 39 73.6 677 1 YOD7_MYCTU HYPOTHETICAL 69.7 KD P 1.65e+01
31 39 73.6 1265 1 RPOD_CYAPA DNA-DIRECTED RNA POLYM 1.65e+01
32 39 73.6 1377 1 RPOC_BORBU DNA-DIRECTED RNA POLYM 1.65e+01
33 38 71.7 182 1 CYPC_YEAST PEPTIDYL-PROLYL CIS-TR 2.93e+01
34 38 71.7 264 1 YNV6_YEAST HYPOTHETICAL 30.6 KD P 2.93e+01
35 38 71.7 361 1 HA1B_RABIT RLA CLASS 1 HISTOCOMPA 2.93e+01
36 38 71.7 361 1 HA1B_RABIT RLA CLASS 1 HISTOCOMPA 2.93e+01
37 38 71.7 375 1 NUEM_NEUCR NADH-UBIQUINONE OXIDOR 2.93e+01
38 38 71.7 393 1 KGCY_NERDI GLYCOCYAMINE KINASE (E 2.93e+01
39 38 71.7 485 1 SAHH_PHASS ADENOSYLHOMOCYSTEINASE 2.93e+01
40 38 71.7 540 1 PYRG_METUA CTP SYNTHASE (EC 6.3.4 2.93e+01
41 38 71.7 684 1 CNG1_MOUSE CGMP-GATED CATION CHAN 2.93e+01
42 38 71.7 707 1 DCOR_LEIDO ORNITHINE DECARBOXYLAS 2.93e+01
43 38 71.7 1045 1 PRIS_SERMA EXTRACELLULAR SERINE P 2.93e+01
44 38 71.7 1073 1 POL_HV2DI POLYPROTEIN [CONTA 2.93e+01
45 38 71.7 1210 1 ICEN_PSEFL ICE NUCLEATION PROTEIN 2.93e+01

ALIGNMENTS

RESULT 1
ID PHS1_SOLTU STANDARD; PRT; 966 AA.
AC P04045;
DT 01-NOV-1986 (REL. 03, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ALPHA-GLUCAN PHOSPHORYLASE, L ISOZYME 1 PRECURSOR (EC 2.4.1.1)
DE (STARCH PHOSPHORYLASE L-1).
OS SOLANUM TUBEROSUM (POTATO).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANACEAE; SOLANUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90110071.
RA NAKANO K., MORI H., FUKUI T.;
RT "Molecular cloning of cDNA encoding potato amyloplast alpha-glucan
phosphorylase and the structure of its transit peptide.";
RL J. BIOCHEM. 106:691-695(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, KENNEBEC;
RX MEDLINE; 91117174.
RA CAMIRAND A., ST PIERRE B., MARINEAU C., BRISSON N.;
RT "Occurrence of a copia-like transposable element in one of the
introns of the potato starch phosphorylase gene.";
RL MOL. GEN. GENET. 224:33-39(1990).
RN [3]
RP SEQUENCE OF 51-966.
RX MEDLINE; 86250715.
RA NAKANO K., FUKUI T.;
RT "The complete amino acid sequence of potato alpha-glucan
phosphorylase.";
RL J. BIOL. CHEM. 261:8230-8236(1986).
RN [4]
RP SEQUENCE OF 51-131.
RX MEDLINE; 81007008.
RA NAKANO K., FUKUI T., MATSUBARA H.;
RT "Structural basis for the difference of the regulatory properties
between potato and rabbit muscle phosphorylases. The NH2-terminal
sequence of the potato enzyme.";
RL J. BIOL. CHEM. 255:9255-9261(1980).
CC -[- FUNCTION: PHOSPHORYLASE IS AN IMPORTANT ALLOSTERIC ENZYME IN
CARBOHYDRATE METABOLISM. ENZYMES FROM DIFFERENT SOURCES DIFFER IN
THEIR REGULATORY MECHANISMS AND IN THEIR NATURAL SUBSTRATES.
CC HOWEVER, ALL KNOWN PHOSPHORYLASES SHARE CATALYTIC AND STRUCTURAL
PROPERTIES.

CC -|- CATALYTIC ACTIVITY: (1,4-ALPHA-D-GLUCOSYL)(N) + ORTHOPHOSPHATE =
CC (1,4-ALPHA-D-GLUCOSYL)(N-1) + ALPHA-D-GLUCOSE 1-PHOSPHATE.
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -|- SUBCELLULAR LOCATION: CHLOROPLAST; AMYLOPLAST.
CC -|- TISSUE SPECIFICITY: TUBER.
CC -|- SIMILARITY: BELONGS TO THE GLYCOGEN PHOSPHORYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D00520; D1000861; -
CC EMBL: X52385; G21579; -
CC PIR: JU0130; PHPOAG.
CC PROSITE: PS00102; PHOSPHORYLASE; 1.
CC PFAM: PF00343; phosphorylase; 2.
CC HSP: P06738; IYGP.
CC TRANSFERASE: GLYCOSYLTRANSFERASE; CARBOHYDRATE METABOLISM;
CC ALLOSTERIC ENZYME: PYRIDOXAL PHOSPHATE; TRANSIT PEPTIDE; CHLOROPLAST;
CC AMYLOPLAST; MULTIGENE FAMILY.
CC TRANSIT 1 50 CHLOROPLAST.
CC CHAIN 51 966 ALPHA-GLUCAN PHOSPHORYLASE, L ISOZYME 1.
CC BINDING 812 812 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC CONFLICT 159 159 A -> D (IN REF. 2).
CC SEQUENCE 966 AA; 109505 MW; C98DF124 CRC32;
CC
Query Match 90.6%; Score 48; DB 1; Length 966;
Best Local Similarity 77.8%; Pred. No. 5.45e-02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 62 EQGESDLS 70
QY 237 DOGESLS 245
:|||||
RESULT 2
ID YGGJ_ERWCH STANDARD; PRT; 93 AA.
AC P37995;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN IN NUC 3' REGION (FRAGMENT).
OS ERWINIA CHRYSANTHEMI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN:3937;
CC MEDLINE; 9323747.
CC MOULARD M., CONDEME G., ROBERT-BAUDOUY J.;
CC "Characterization of the nuc gene coding for a nuclease of the
CC phytopathogenic bacteria Erwinia chrysanthemi.";
CC MOL. MICROBIOL. 8:685-695(1993).
CC -|- SIMILARITY: STRONG, TO E.COLI YGGJ AND H.INFLUENZAE HI0303.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X73255; G403003; -
CC PIR: S42743; S42743.
CC HYPOTHETICAL PROTEIN.
CC NON_TER 93 93
CC SEQUENCE 93 AA; 10283 MW; 45AEB6D4 CRC32;
CC

Query Match 77.4%; Score 41; DB 1; Length 93;
Best Local Similarity 75.0%; Pred. No. 5.01e-00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 15 QGGETDLS 22
QY 238 QGGSLS 245
:|||||
RESULT 3
ID CREB_CHLVR STANDARD; PRT; 249 AA.
AC P51984;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CAMP RESPONSE ELEMENT BINDING PROTEIN.
CC CREB.
CC CHLOROHYDRA VIRIDISSIMA (HYDRA).
CC EUKARYOTA; METAZOA; CNIDARIA; HYDROZOA; ANTHOMEDUSAE;
CC HYDRIDAE; CHLOROHYDRA.
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE; 95262565.
CC GALLIOT B., WELSCHOF M., SCHUCKERT O., HOFFMEISTER S., SCHALLER H.C.;
CC "The camp response element binding protein is involved in hydra
CC regeneration".
CC DEVELOPMENT 121:1205-1209(1995).
CC -|- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE),
CC SEQUENCE PRESENT IN MANY VIRAL AND CELLULAR PROMOTERS. COULD
CC REGULATE THE TRANSCRIPTIONAL ACTIVITY OF GENES INVOLVED IN
CC REGENERATION PROCESSES.
CC -|- SUBUNIT: BINDS DNA AS A DIMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: NUCLEAR
CC -|- SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X83873; G633191; -
CC DR PROSITE: PS00036; BZIP_BASIC; 1.
CC DR PFAM: PF00170; bzip; 1.
CC DR HSP: P03089; IYSA.
CC KW TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN.
CC FT DNA_BIND 192 213 BASIC MOTIF (BY SIMILARITY).
CC FT DOMAIN 219 240 LEUCINE-ZIPPER (BY SIMILARITY).
CC SEQUENCE 249 AA; 27950 MW; F82954E2 CRC32;
CC
Query Match 77.4%; Score 41; DB 1; Length 249;
Best Local Similarity 75.0%; Pred. No. 5.01e-00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 132 QGGSLS 139
QY 238 QGGSLS 245
:|||||
RESULT 4
ID ASPG_MOUSE STANDARD; PRT; 346 AA.
AC Q64191;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE N4-(BETA-N-ACETYLGLUCOSAMINYL)-L-ASPARAGINASE PRECURSOR (EC 3.5.1.26)
DE (GLYCOSYLASPARAGINASE) (ASPARTYLGLUCOSAMINIDASE) (N4-(N-ACETYL-BETA-
DE GLUCOSAMINYL)-L-ASPARAGINE AMIDASE) (AGA).
CC AGA.
CC MUS MUSCULUS (MOUSE).
CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
CC


```

RESULT 7
ID VIL_HPV41 STANDARD; PRT; 534 AA.
AC P27557;
DT 01-AUG-1992 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MAJOR CAPSID PROTEIN L1.
GN L1.
OS HUMAN PAPILLOMAVIRUS TYPE 41.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91253264.
RA HIRT L., HIRSCH-BEHNAM A., DE VILLIERS E.M.;
RT "Nucleotide sequence of human papillomavirus (HPV) type 41: an
RT unusual HPV type without a typical E2 binding site consensus
RT sequence.";
RL VIRUS RES. 18:179-190(1990).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X56147; G60950; ALT_INIT.
DR PIR; H43550; P1WL41.
DR PFAM; PF00500; late_protein_l1; 1.
DR COAT PROTEIN; LATE PROTEIN.
SQ SEQUENCE 534 AA; 60949 MW; D58B8EFC CRC32;
-----
Query Match 77.4%; Score 41; DB 1; Length 534;
Best Local Similarity 66.7%; Pred. No. 5.01e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 163 DOGSDSRLS 171
|||:| |
QY 237 DOGGESSL 245

RESULT 8
ID YH04 YEAST STANDARD; PRT; 796 AA.
AC P38888;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 91.2 KD PROTEIN IN RPS4B-SCH9 INTERGENIC REGION.
GN YHR204W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN; S288C / AB972;
RX MEDLINE; 94378003.
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,
RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
RA NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,
RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
RA VAUDIN M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL SCIENCE 265:2077-2082(1994).
-----
CC -1- SIMILARITY: LOW, TO FAMILY 47 OF GLYCOSYL HYDROLASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; U00030; G458945; -.
DR PIR; S46693; S46693.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 796 AA; 91245 MW; 291F8C79 CRC32;
-----
Query Match 77.4%; Score 41; DB 1; Length 796;
Best Local Similarity 55.6%; Pred. No. 5.01e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 534 EQAGETLS 542
:|:|:|:|
QY 237 DOGGESSL 245

RESULT 9
ID FASD_ECOLI STANDARD; PRT; 835 AA.
AC P46000;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE USHER PROTEIN FASD PRECURSOR.
GN FASD.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN; 987;
RX MEDLINE; 94148769.
RA SCHIFFERLI D.M., ALRUTZ M.A.;
RT "Permissive linker insertion sites in the outer membrane protein of
RT 987P fimbriae of Escherichia coli.";
RL J. BACTERIOL. 176:1099-1110(1994).
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE 987P
CC FIMBRIAE SUBUNITS ACROSS THE OUTER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L22659; G437336; -.
DR EMBL; U50547; G1381551; -.
DR PROSITE; PS01151; FIMBRIAL_USHER; 1.
DR PFAM; PF00577; Usher; 1.
KW OUTER MEMBRANE; TRANSMEMBRANE; FIMBRIA; TRANSPORT; SIGNAL.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 835 OUTER MEMBRANE USHER PROTEIN FASD.
FT DISULFID 810 834 POTENTIAL.
SQ SEQUENCE 835 AA; 92354 MW; D8FBD31 CRC32;
-----
Query Match 77.4%; Score 41; DB 1; Length 835;
Best Local Similarity 85.7%; Pred. No. 5.01e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 689 QGGETSL 695
|||:| |
QY 238 QGGESSL 244

RESULT 10
ID RPWL_SCHPO STANDARD; PRT; 1120 AA.
AC O13993;

```


DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE DNA-DIRECTED RNA POLYMERASE MITOCHONDRIAL PRECURSOR (EC 2.7.7.6).
 GN SPAC26H5.12.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
 OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 OC SCHIZOSACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
 RL SUBMITTED (SEP-1997) TO ENBL/GENBANK/DBJ DATA BANKS
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
 CC RNA(N).
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
 CC -!- SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z99126; E343565;
 DR PROSITE: PS00900; RNA_POL_PHASE.1; 1.
 DR PROSITE: PS00489; RNA_POL_PHASE.2; 1.
 DR PFAM: PF00940; RNA_pol; 1.
 KW TRANSFERASE; DNA-DIRECTED RNA POLYMERASE; TRANSCRIPTION;
 KW MITOCHONDRION; TRANSIT PEPTIDE.
 FT TRANSIT 1 ? MITOCHONDRION
 FT CHAIN ? 1120 DNA-DIRECTED RNA POLYMERASE.
 FT ACT_SITE 787 787 BY SIMILARITY.
 FT ACT_SITE 856 856 BY SIMILARITY.
 FT ACT_SITE 1027 1027 BY SIMILARITY.
 SQ SEQUENCE 1120 AA; 127268 MW; 26525536 CRC32;

 Query Match 77.4%; Score 41; DB 1; Length 1120;
 Best Local Similarity 75.0%; Pred. No. 5.01e+00;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 Db 193 DONGDSSL 200
 Qy 237 DQGESL 244

 RESULT 11
 ID SPCB HUMAN STANDARD; PRT: 2137 AA.
 AC P11277; Q15519; Q15510;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE SPECTRIN BETA CHAIN, ERYTHROCYTE.
 GN SPTB1 OR SPTB.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 90307707.
 RA WINKELMANN J.C., CHANG J.G., TSE W.T., SCARPA A.L., MARCHESI V.T.,
 RA FORGET B.G.;
 RT "Full-length sequence of the cDNA for human erythroid beta-spectrin";
 RL J. BIOL. CHEM. 265:11827-11832(1990).
 RN [2]
 RP SEQUENCE OF 1055-2137 FROM N.A. (MUSCLE FORM).

RC TISSUE=SKELETAL MUSCLE;
 RX MEDLINE: 91056094
 RA WINKELMANN J.C., COSTA F.F., LINZIE B.L., FORGET B.G.;
 RT "Beta spectrin in human skeletal muscle. Tissue-specific differential
 RT processing of 3' beta spectrin pre-mRNA generates a beta spectrin
 RT isoform with a unique carboxyl terminus";
 RL J. BIOL. CHEM. 265:20449-20454(1990).
 RN [3]
 RP SEQUENCE OF 2002-2137 FROM N.A.
 RX MEDLINE: 91332035.
 RA GALLAGHER P.G., TSE W.T., COSTA F., SCARPA A., BOIVIN P., DELAUNAY J.,
 RA FORGET B.G.;
 RT "A splice site mutation of the beta-spectrin gene causing exon
 RT skipping in hereditary elliptocytosis associated with a truncated
 RT beta-spectrin chain";
 RL J. BIOL. CHEM. 266:15154-15159(1991).
 RN [4]
 RP SEQUENCE OF 928-1755 FROM N.A.
 RX MEDLINE: 91007291.
 RA YOON S.H., KENTROS C.G., PRCHAL J.T.;
 RT "Identification of an unusual deletion within homologous repeats of
 RT human reticulocyte beta-spectrin and probable peptide polymorphism";
 RL GENE 91:297-302(1990).
 RN [5]
 RP SEQUENCE OF 1334-1432 AND 1909-2137 FROM N.A.
 RX MEDLINE: 88269838.
 RA WINKELMANN J.C., LETO T.L., WATKINS P.C., EDDY R., SHOWS T.B.,
 RA LINNENBACH A.J., SAHR K.E., KATHURIA N., MARCHESI V.T., FORGET B.G.;
 RT "Molecular cloning of the cDNA for human erythrocyte beta-spectrin";
 RL BLOOD 72:328-334(1988).
 RN [6]
 RP SEQUENCE OF 1209-1482 FROM N.A.
 RX MEDLINE: 88041127.
 RA PRCHAL J.T., MORLEY B.J., YOON S.-H., COETZER T.L., PALEK J.,
 RA CONBOY J.G., KAN Y.W.;
 RT "Isolation and characterization of cDNA clones for human erythrocyte
 RT beta-spectrin";
 RL PROC. NATL. ACAD. SCI. U.S.A. 84:7468-7472(1987).
 RN [7]
 RP DOMAINS.
 RX MEDLINE: 84295638.
 RA SPEICHER D.W., MARCHESI V.T.;
 RT "Erythrocyte spectrin is comprised of many homologous triple helical
 RT segments";
 RL NATURE 311:177-180(1984).
 RN [8]
 RP REVIEW ON VARIANTS.
 RX MEDLINE: 97001215.
 RA MAILLET P., ALLOISIO N., MORLE L., DELAUNAY J.;
 RT "Spectrin mutations in hereditary elliptocytosis and hereditary
 RT spherocytosis";
 RL HUM. MUTAT. 8:97-107(1996).
 RN [9]
 RP VARIANT HE CAGLIARI.
 RX MEDLINE: 94043025.
 RA SAHR K.E., COETZER T.L., MOY L.S., DERICK L.H., CHISHTI A.H.,
 RA JAROLIM P., LORENZO F., DEL GIUDICE E.M., IOLASCON A., GALLIANELLO R.,
 RA CAO A., DELAUNAY J., LIU S.-C., PALEK J.;
 RT "Spectrin Cagliari: an Ala->Gly substitution in helix 1 of beta
 RT spectrin repeat 17 that severely disrupts the structure and self-
 RT association of the erythrocyte spectrin heterodimer";
 RL J. BIOL. CHEM. 268:22656-22662(1993).
 RN [10]
 RP VARIANT HE KISSIMMEE.
 RX MEDLINE: 93352802.
 RA BECKER P.S., TSE W.T., LUX S.E., FORGET B.G.;
 RT "Beta spectrin Kissimmee: a spectrin variant associated with
 RT autosomal dominant hereditary spherocytosis and defective binding to
 RT protein 4.1";
 RL J. CLIN. INVEST. 92:612-616(1993).
 RN [11]
 RP VARIANT HE PROVIDENCE PRO-2019.
 RX MEDLINE: 95190014.

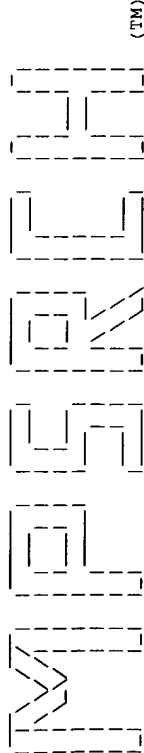
RA GALLAGHER P.G., WEED S.A., TSE W.T., BENOIT L., MORROW J.S.,
RA MARCHESI S.L., MOHANDAS N., FORGET B.G.;
RT "Recurrent fatal hydrops fetalis associated with a nucleotide
RT substitution in the erythrocyte beta-spectrin gene.";
RL J. CLIN. INVEST. 95:1174-1182(1995).
RN [12]
RN VARIANTS HE VAL-2023 AND ARG-2024.
RP MEDLINE; 94289716.
RX PARQUET N., DEVAUX I., BOULANGER L., GALAND C., BOIVIN P.,
RA LECOMTE M.-C., DHERMY D., GARBARZ M.;
RT "Identification of three novel spectrin alpha I/74 mutations in
RT hereditary elliptocytosis: further support for a triple-stranded
RT folding unit model of the spectrin heterodimer contact site.";
RL BLOOD 84:303-308(1994).
RN [13]
RN VARIANT HE PRO-2053.
RP MEDLINE; 90369011.
RX TSE W.T., LECOMTE M.-C., COSTA F.F., GARBARZ M., FEO C., BOIVIN P.,
RA DHERMY D., FORGET B.G.;
RT "Point mutation in the beta-spectrin gene associated with alpha I/74
RT hereditary elliptocytosis. Implications for the mechanism of spectrin
RT dimer self-association.";
RL J. CLIN. INVEST. 86:909-916(1990).
RN [14]
RN FUNCTION: SPECTRIN IS THE MAJOR CONSTITUENT OF THE CYTOSKELETAL
CC NETWORK UNDERLYING THE ERYTHROCYTE PLASMA MEMBRANE. IT ASSOCIATES
CC WITH BAND 4.1 AND ACTIN TO FORM THE CYTOSKELETAL SUPERSTRUCTURE OF
CC THE ERYTHROCYTE PLASMA MEMBRANE.
CC -!- SUBUNIT: COMPOSED OF NONHOMOLOGOUS CHAINS, ALPHA AND BETA, WHICH
CC AGGREGATE TO FORM DIMERS, TETRAMERS, AND HIGHER POLYMERS.
CC -!- DISORDER: HEREDITARY ELLIPTOCYTOSIS (HE) IS A HETEROGENEOUS
CC DISEASE CHARACTERIZED BY VARIABLE HEMOLYTIC ANEMIA AND ELLIPTICAL
CC RED CELL SHAPE. SEVERAL ABNORMALITIES IN THE MEMBRANE SKELETON
CC HAVE BEEN IDENTIFIED IN HE, INCLUDING A NUMBER THAT HAVE BEEN
CC LOCALIZED TO SPECTRIN. ALSO INVOLVED IN HEREDITARY SPHEROCYTOSIS
CC (HS).
CC -!- THIS COMPLEX IS ANCHORED TO THE CYTOPLASMIC FACE OF THE PLASMA
CC MEMBRANE VIA ANOTHER PROTEIN, ANKYRIN, WHICH BINDS TO BETA-
CC SPECTRIN AND MEDIATES THE BINDING OF THE WHOLE COMPLEX TO A
CC TRANSMEMBRANE PROTEIN BAND 3. THE INTERACTION OF ERYTHROCYTE
CC SPECTRIN WITH OTHER PROTEINS THROUGH SPECIFIC BINDING DOMAINS
CC LEAD TO THE FORMATION OF AN EXTENSIVE SUBPLASMALEMAL MESHWORK
CC WHICH IS THOUGHT TO BE RESPONSIBLE FOR THE MAINTENANCE OF THE
CC BICONCAVE SHAPE OF HUMAN ERYTHROCYTES, FOR THE REGULATION OF THE
CC PLASMA MEMBRANE COMPONENTS AND FOR THE MAINTENANCE OF THE LIPID
CC ASYMMETRY OF THE PLASMA MEMBRANE.
CC -!- ALTERNATIVE PRODUCTS: A MUSCLE-SPECIFIC FORM IS PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J05500; G338440; -
CC EMBL; J05500; G338441; -
CC EMBL; M57948; -; NOT_ANNOTATED_CDS.
CC EMBL; M37884; G338330; -
CC EMBL; M37885; G338332; -
CC EMBL; M18054; G338334; -
CC EMBL; X59510; G29499; -
CC EMBL; X59511; G29497; -
CC PIR; A37064; SCHUB.
CC SWISS-2DPAGE; P11277; HUMAN.
CC MIM; 182870; -
CC MIM; 130600; -
CC PROSITE; PS00019; ACTININ_1; 1.
CC PROSITE; PS00020; ACTININ_2; 1.
CC PFAM; PF00307; actinin-binding; 1.
CC PFAM; PF00435; spectrin; 17.

DR HSSP; Q01082; IAA2.
KW CYTOSKELETON; MEMBRANE; ERYTHROCYTE; REPEAT; ACTIN-BINDING;
KW CAPING PROTEIN; DISEASE MUTATION; ELLIPTOCYTOSIS;
KW ALTERNATIVE SPLICING.
FT DOMAIN 1 272 ACTIN-BINDING.
FT REPEAT 273 384 1.
FT REPEAT 385 498 2.
FT REPEAT 499 607 3.
FT REPEAT 608 713 4.
FT REPEAT 714 818 5.
FT REPEAT 819 924 6.
FT REPEAT 925 1031 7.
FT REPEAT 1032 1138 8.
FT REPEAT 1139 1244 9.
FT REPEAT 1245 1349 10.
FT REPEAT 1350 1455 11.
FT REPEAT 1456 1554 12.
FT REPEAT 1555 1660 13.
FT REPEAT 1661 1767 14.
FT REPEAT 1768 1873 15.
FT REPEAT 1874 1979 16.
FT REPEAT 1980 2085 17.
FT VARIANT 202 202 W -> R (IN KISSIMMEE; HS).
FT VARIANT 439 439 N -> S.
FT VARIANT 1151 1151 D -> N.
FT VARIANT 1374 1374 H -> R.
FT VARIANT 1403 1403 R -> Q.
FT VARIANT 2018 2018 A -> G (IN CAGLIARY; HE).
FT VARIANT 2019 2019 S -> P (IN PROVIDENCE; HE).
FT VARIANT 2023 2023 A -> V (IN PARIS; HE).
FT VARIANT 2024 2024 W -> R (IN LINGUERE; HE).
FT VARIANT 2025 2025 L -> R (IN BUFFALO; HE).
FT VARIANT 2053 2053 A -> P (IN KAYES; HE).
FT VARSPLIC 2116 2137 VSLWSRLSSSHESLOPEFSHY -> GEEEGTWPNLQPPP
PPGQHKQKQKSTGDRPTTEPLFKVLDTPLSEGEDEPATLPA
...
Note: remainder of annotations omitted.
Query Match 77.4%; Score 41; DB 1; Length 2137;
Best Local Similarity 44.4%; Pred. No. 5.01e-00; Mismatches 4; Indels 0; Gaps 0;
Matches 4; Conservative 4; Disabled Protein.
Db 1447 ERGGDADLS 1455
QY 237 DOGGESSLS 245
:::|::|
RESULT 12
ID DAB_DROME STANDARD; PRT; 2411 AA.
AC P98081;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE DISABLED PROTEIN.
GN DAB.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93194063.
RA GERTLER F.B., HILL K.K., CLARK M.J., HOFFMANN F.M.;
RT "Dosage-sensitive modifiers of Drosophila abl tyrosine kinase
RT function: prospero, a regulator of axonal outgrowth, and disabled, a
RT novel tyrosine kinase substrate.";
RL GENES DEV. 7:441-453(1993).
CC -!- FUNCTION: TOGETHER WITH ABL INVOLVED IN EMBRYONIC NEURAL
CC DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- TISSUE SPECIFICITY: CNS AXONS AND BODY WALL MUSCLES.
CC -!- DEVELOPMENTAL STAGE: EMBRYONIC AXONEMESIS.
CC -!- PTM: PROBABLY PHOSPHORYLATED BY THE ABL TYROSINE KINASE.

```
CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS ARE PRODUCED FROM THE SAME GENE
CC BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L08845; G1498252; ALT_SEQ.
CC PIR: A46299; A46299.
CC FLYBASE: FBgn0000414; Dab.
CC PROSITE: PS01179; PID; 1.
CC PFAM: PF00640; PID; 1.
CC ALTERNATIVE SPLICING; PHOSPHORYLATION.
CC DOMAIN 46 196 PID.
CC VARSPLIC 462 673 MISSING (IN SHORTER FORM).
CC REPEAT 1689 1801 REPEAT-RICH REGION.
CC REPEAT 1689 1700 ALTERNATE ARG AND ACIDIC RESIDUE.
CC REPEAT 1740 1750 ALTERNATE ARG AND ACIDIC RESIDUE.
CC REPEAT 1791 1801 ALTERNATE ARG AND ACIDIC RESIDUE.
CC MOD_RES 111 111 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC MOD_RES 482 482 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC MOD_RES 1662 1662 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC MOD_RES 1667 1667 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC MOD_RES 1701 1701 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC MOD_RES 1704 1704 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC MOD_RES 1713 1713 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC MOD_RES 1739 1739 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC MOD_RES 1826 1826 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC MOD_RES 1961 1961 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC SEQUENCE 2411 AA; 264046 MW; 4EC2EDA7 CRC32;
CC -----
Query Match 77.4%; Score 41; DB 1; Length 2411;
Best Local Similarity 87.5%; Pred. No. 5.01e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CC -----
Db 1306 DAGGESSL 1313
QY 237 DOGGESSL 244
: |||||
RESULT 13
ID RRPV_VSVJM STANDARD; PRT; 274 AA.
AC P04878;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DE RNA POLYMERASE ALPHA SUBUNIT (EC 2.7.7.48) (NONSTRUCTURAL
DE PHOSPHOPROTEIN).
GN NS.
OS VESICULAR STOMATITIS VIRUS (SEROTYPE NEW JERSEY / STRAIN MISSOURI).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC RIADOVIRIDAE; VESICULOVIRUS.
[1]
SEQUENCE FROM N.A.
RAE B.P., ELLIOTT R.M.;
MEDLINE: 86253157.
RT "Conservation of potential phosphorylation sites in the NS proteins
RT of the New Jersey and Indiana serotypes of vesicular stomatitis
RT virus."
RL J. GEN. VIROL. 67:1351-1360(1986).
CC -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
CC POLYMERASE. IT MAY FUNCTION IN TEMPLATE BINDING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04063; G61849; -.
CC PIR: A29143; MNVNVJ.
CC PFAM: PF00922; Phosphoprotein; 1.
CC TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; PHOSPHORYLATION;
CC NONSTRUCTURAL PROTEIN.
CC SEQUENCE 274 AA; 31300 MW; CDA00585 CRC32;
CC -----
Query Match 75.5%; Score 40; DB 1; Length 274;
Best Local Similarity 66.7%; Pred. No. 9.14e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC -----
Db 110 ENGGEKSL 118
QY 237 DOGGESSL 245
: ||| |||
RESULT 15
ID YDEM_ECOLI STANDARD; PRT; 385 AA.
AC P76134; P77755;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE HYPOTHETICAL 44.5 KD PROTEIN IN POOL-HIPA INTERGENIC REGION.
GN YDEM.
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04063; G61849; -.
CC PIR: A29143; MNVNVJ.
CC PFAM: PF00922; Phosphoprotein; 1.
CC TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; PHOSPHORYLATION;
CC NONSTRUCTURAL PROTEIN.
CC SEQUENCE 274 AA; 31300 MW; CDA00585 CRC32;
CC -----
Query Match 75.5%; Score 40; DB 1; Length 274;
Best Local Similarity 66.7%; Pred. No. 9.14e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC -----
Db 110 ENGGEKSL 118
QY 237 DOGGESSL 245
: ||| |||
RESULT 14
ID RRPV_VSVJO STANDARD; PRT; 274 AA.
AC P04877;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 13-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE ALPHA SUBUNIT (EC 2.7.7.48) (NONSTRUCTURAL
DE PHOSPHOPROTEIN).
GN NS.
OS VESICULAR STOMATITIS VIRUS (SEROTYPE NEW JERSEY / STRAIN OGDEN).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC RIADOVIRIDAE; VESICULOVIRUS.
[1]
SEQUENCE FROM N.A.
RA GILL D.S., BANERJEE A.K.;
MEDLINE: 85237710.
RX "Vesicular stomatitis virus NS proteins: structural similarity
RT without extensive sequence homology."
RL J. VIROL. 55:60-66(1985).
CC -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
CC POLYMERASE. IT MAY FUNCTION IN TEMPLATE BINDING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: K03387; G336048; -.
CC PIR: A04115; MNVNVJ.
CC PFAM: PF00922; Phosphoprotein; 1.
CC TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; PHOSPHORYLATION;
CC NONSTRUCTURAL PROTEIN.
CC SEQUENCE 274 AA; 31406 MW; 53219599 CRC32;
CC -----
Query Match 75.5%; Score 40; DB 1; Length 274;
Best Local Similarity 66.7%; Pred. No. 9.14e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC -----
Db 110 ENGGEKSL 118
QY 237 DOGGESSL 245
: ||| |||
RESULT 15
ID YDEM_ECOLI STANDARD; PRT; 385 AA.
AC P76134; P77755;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE HYPOTHETICAL 44.5 KD PROTEIN IN POOL-HIPA INTERGENIC REGION.
GN YDEM.
```

•



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:43:35 1999; MasPar time 10.53 Seconds
Tabular output not generated. 20.203 Million cell updates/sec

Title: >US-09-041-236-2
Description: (247-256) from US09041236.pep (18 of 45)
Perfect Score: 80
Sequence: 1 SKWNTFLKAM 10

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21265608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 18.369; Variance 63.609; scale 0.289

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	62	77.5	W63748	Human semaphorin.	2.38e+01
2	61	76.3	R74175	Human collapsin.	2.98e+01
3	61	76.3	R71380	Human semaphorin III	2.98e+01
4	57	71.3	W51314	Human semaphorin W.	7.20e+01
5	57	71.3	W51313	Rat semaphorin W.	7.20e+01
6	57	71.3	W17658	Mouse CD100 antigen.	7.20e+01
7	57	71.3	W58540	Human semaphorin.	7.20e+01
8	57	71.3	W17657	Human CD100 antigen.	7.20e+01
9	54	67.5	R26167	85-C.	1.38e+02
10	54	67.5	R71381	Vaccinia virus semaph	1.38e+02
11	54	67.5	R71384	Tribolium semaphorin	1.38e+02
12	53	66.3	W62223	HAP4 protein.	1.71e+02
13	53	66.3	W71382	Drosophila semaphorin	1.71e+02
14	52	65.0	15 34	M. tuberculosis 30 kD	2.11e+02
15	52	65.0	15 34	M. tuberculosis 32A k	2.11e+02
16	52	65.0	15 34	M. tuberculosis 32A k	2.11e+02

17	52	65.0	15 34	W75606	M. tuberculosis 30 kD	2.11e+02
18	52	65.0	26 33	W62325	Bactolysin - antimicr	2.11e+02
19	52	65.0	26 33	W62327	Bactolysin - antimicr	2.11e+02
20	52	65.0	26 33	W07453	Antimicrobial cationi	2.11e+02
21	52	65.0	26 33	W62332	Bactolysin - antimicr	2.11e+02
22	52	65.0	26 33	W07455	Antimicrobial cationi	2.11e+02
23	52	65.0	26 33	W62328	Bactolysin - antimicr	2.11e+02
24	52	65.0	26 33	W07456	Antimicrobial cationi	2.11e+02
25	52	65.0	26 23	W07460	Antimicrobial cationi	2.11e+02
26	52	65.0	36 1	P91330	Amino acid sequence o	2.11e+02
27	52	65.0	137 23	W07857	(DSM 10100) human pap	2.11e+02
28	52	65.0	325 22	W18164	Mycobacterium tubercu	2.11e+02
29	52	65.0	325 2	R08099	Mycobacterium deriv	2.11e+02
30	52	65.0	325 2	R10486	Tuberculin active pro	2.11e+02
31	52	65.0	325 30	W45776	Amino acid sequence o	2.11e+02
32	52	65.0	330 30	W45777	Mycobacterium tubercu	2.11e+02
33	52	65.0	338 2	R11296	Amino acid sequence o	2.11e+02
34	52	65.0	338 2	W18165	Recombinant M. tubercu	2.11e+02
35	52	65.0	338 34	W63033	Mycobacterium tubercu	2.11e+02
36	52	65.0	353 2	R11295	Recombinant M. tubercu	2.11e+02
37	52	65.0	403 37	W72943	Mycobacterium tubercu	2.11e+02
38	52	65.0	404 37	W72942	Mycobacterium tubercu	2.11e+02
39	52	65.0	730 13	R71379	Grasshopper semaphori	2.11e+02
40	52	65.0	402 24	W13497	Plant plastid phospho	2.61e+02
41	51	63.8	673 39	W89587	Sphingomonas capsulat	2.61e+02
42	51	63.8	1107 18	W03132	Human DNA polymerase	2.61e+02
43	51	63.8	26 33	W62336	Bactolysin - antimicr	3.22e+02
44	50	62.5	26 23	W07464	Antimicrobial cationi	3.22e+02
45	50	62.5	26 23	W07464	Antimicrobial cationi	3.22e+02

ALIGNMENTS

RESULT 1
ID W63748 standard; Protein; 775 AA.
AC W63748;
DT 01-OCT-1998 (first entry)
DE Human semaphorin.
KW Human; semaphorin; diagnosis; nerve disease; nerve growth inhibitor;
KW neurological disease; atopic skin inflammation; autoimmune disease;
KW pain.
OS Homo sapiens.
PN W09822504-A1.
PD 28-MAY-1998.
PF 12-NOV-1997; J04111.
PR 15-NOV-1996; JP-321068.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Furuyama T, Inagaki S;
DR WPI: 98-312416/27.
DR N-PSDB; V35367.
PT Gene encoding new semaphorin nerve growth inhibitor - useful in
PT diagnosis, treatment and study of neurological diseases
PS Claim 1; Page 33-37; 49pp; Japanese.
CC The present sequence represents human semaphorin, a nerve growth
CC inhibitor. The semaphorin protein, and gene encoding the protein,
CC and their derivatives, are used in the diagnosis, treatment and
CC study of neurological disorders such as atopic skin inflammation,
CC autoimmune diseases and pain.
SQ Sequence 775 AA;

Query Match 77.5%; Score 62; DB 33; Length 775;
Best Local Similarity 77.8%; Pred. No. 2.38e+01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 282 nkwtstflka 290
Qy 247 SKWNTFLKA 255

RESULT 2

ID R74175 standard; Protein; 477 AA.
AC R74175;
DT 01-NOV-1995 (first entry)

DE Human collapsin.
 KW Collapsin; antibody; therapy.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT binding_site 9..19 "antibody binding site"
 FT binding_site 51..65 "antibody binding site"
 FT US5416197-A.
 PN 16-MAY-1995.
 PD 15-OCT-1993; 136922.
 PR 15-OCT-1993; US-136922.
 PA (OYPE-) UNIV PENNSYLVANIA.
 PI Luo Y, Raper JA;
 DR WPI: 95-193478/25.
 DR N-PSDB: Q92331.
 PT New antibody to human collapsin - used to inhibit the activity of
 PT collapsin, to induce neurite out-growth and to treat individuals with
 PT nerve damage.
 PS Claim 2: Columns 15-18; lipp; English.
 CC An antibody capable of specifically binding at least a portion of
 CC the collapsin protein can be used to purify human collapsin and
 CC to inhibit the activity of the protein. It can be used to induce
 CC neurite outgrowth by neuronal cells and to treat individuals
 CC suffering from nerve damage.
 SQ Sequence 477 AA;

Query Match 76.3%; Score 61; DB 13; Length 477;
 Best Local Similarity 77.8%; Pred. No. 2.98e+01;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 49 nkwtfflka 57
 :|||
 QY 247 SKWNTFLKA 255

RESULT 3
 ID R71380 standard; Protein; 771 AA.
 AC R71380;
 DT 21-NOV-1995 (first entry)
 DE Human semaphorin III protein.
 KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
 KW variola major virus; smallpox; semaphorin receptor binding activity;
 KW modulation; nerve cell growth; immune response; viral pathogenesis;
 KW neurological disease; neuro-regeneration; oncological infection.
 OS Homo sapiens.
 PN W09507706-A.
 PD 23-MAR-1995.
 PF 13-SEP-1994; U10151.
 PR 13-SEP-1993; US-121713.
 PA (REGC) UNIV CALIFORNIA.
 PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
 PI O'Connor T;
 DR WPI: 95-131177/17.
 DR N-PSDB: Q87442.
 PT New class of semaphorin peptide(s) and polypeptide(s) - are
 PT potent modulators of nerve cell growth and regeneration
 PS Example 2: Page 60-63; 101pp; English.
 CC The sequence of the human semaphorin III protein. The proteins
 CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
 CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
 CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
 CC virus semaphorin IV (Q87447) genes were used to generate a series of
 CC peptides (R70370-R70418), which retain semaphorin receptor binding
 CC activity. The semaphorin derived or semaphorin receptor derived peptides
 CC are potent modulators of nerve cell growth, immune responsiveness and
 CC viral pathogenesis. They can be used in diagnosis and treatment of
 CC neurological disease and neuro-regeneration, immune modulation and
 CC diagnosis and treatment of viral and oncological infection and diseases.
 SQ Sequence 771 AA;

Query Match 76.3%; Score 61; DB 13; Length 771;
 Best Local Similarity 77.8%; Pred. No. 2.98e+01;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 281 nkwtfflka 289
 :|||
 QY 247 SKWNTFLKA 255
 RESULT 4
 ID W51314 standard; Protein; 587 AA.
 AC W51314;
 DT 08-SEP-1998 (first entry)
 DE Human semaphorin W.
 KW Human; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
 KW immunosuppressant; gene therapy; diagnosis; research reagent.
 OS Homo sapiens.
 PN W09815628-A1.
 PD 16-APR-1998.
 PF 03-OCT-1997; J03549.
 PR 09-OCT-1996; JP-287636.
 PA (SUMU) SUMITOMO PHARM CO LTD.
 PI Kikuchi K, Kimura T;
 DR WPI: 98-261015/23.
 DR N-PSDB: V07281, V07282.
 PT Nerve extension inhibitor protein semaphorin W - is useful as
 PT therapeutic drug and diagnostic and research reagent
 PS Example 4: Page 69-72; 90pp; Japanese.
 CC The present sequence represents human semaphorin W. Semaphorin W and
 CC its derivatives are nerve extension inhibitors which are useful as
 CC antiallergic, immunosuppressant and anticancer agents. The DNA
 CC encoding semaphorin W can also be used in gene therapy, e.g. using
 CC a viral vector. The proteins, peptides, DNA and antibodies which
 CC recognise the protein or peptides, can be used as diagnostic or
 CC research reagents. Semaphorin W can be used as a screen for
 CC semaphorin W antagonists with possible therapeutic use.
 SQ Sequence 587 AA;

Query Match 71.3%; Score 57; DB 32; Length 587;
 Best Local Similarity 75.0%; Pred. No. 7.20e+01;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 103 rwtfflka 110
 :|||
 QY 248 KWNTEFLKA 255

RESULT 5
 ID W51313 standard; Protein; 776 AA.
 AC W51313;
 DT 08-SEP-1998 (first entry)
 DE Rat semaphorin W.
 KW Rat; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
 KW immunosuppressant; gene therapy; diagnosis; research reagent.
 OS Rattus norvegicus.
 PN W09815628-A1.
 PD 16-APR-1998.
 PF 03-OCT-1997; J03549.
 PR 09-OCT-1996; JP-287636.
 PA (SUMU) SUMITOMO PHARM CO LTD.
 PI Kikuchi K, Kimura T;
 DR WPI: 98-261015/23.
 DR N-PSDB: V07279, V07280.
 PT Nerve extension inhibitor protein semaphorin W - is useful as
 PT therapeutic drug and diagnostic and research reagent
 PS Claim 1: Page 60-64; 90pp; Japanese.
 CC The present sequence represents rat semaphorin W. Semaphorin W and
 CC its derivatives are nerve extension inhibitors which are useful as
 CC antiallergic, immunosuppressant and anticancer agents. The DNA
 CC encoding semaphorin W can also be used in gene therapy, e.g. using
 CC a viral vector. The proteins, peptides, DNA and antibodies which
 CC recognise the protein or peptides, can be used as diagnostic or
 CC research reagents. Semaphorin W can be used as a screen for
 CC semaphorin W antagonists with possible therapeutic use.
 SQ Sequence 776 AA;

```

Query Match          71.3%; Score 57; DB 32; Length 776;
Best Local Similarity 75.0%; Pred. No. 7.20e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 291 rwtflka 298
   :| |||||
Qy 248 KWNTEFLKA 255

RESULT 6
ID W17658 standard; Protein; 861 AA.
AC W17658;
DE 24-JUL-1997 (first entry)
DE Mouse CD100 antigen.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
KW vaccine.
OS Mus sp.
FH Key Location/Qualifiers
FT peptide 1..41
FT /label= Sig_peptide
FT protein 42..861
FT /label= Mat_protein
FT domain 42..553
FT /label= Semaphorin_domain
FT domain 554..630
FT /label= Ig-like_domain
FT domain 631..732
FT /label= Stalk_domain
FT domain 734..752
FT /label= Transmembrane_domain
FT domain 753..861
FT /label= Cytoplasmic_domain
FT modified_site 807..814
FT /label= Phosphorylation
FT /note= "putative tyrosine phosphorylation site".
PN W09717368-AL.
PD 15-MAY-1997.
PR 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND ) DANA FARBER CANCER INST.
PI Boussioutis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB; T60665.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Example 8; Page 86-89; 135pp; English.
CC Mouse CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60666) isolated from murine T cells. Human CD100 antigen
CC (W17657) has also been identified. CD100 polypeptides and fusion
CC proteins, nucleic acids, and host cells expressing CD100 can be
CC utilised in diagnostic and therapeutic methods involving modulation
CC of B and T cell responses, neuron axonal growth and immune cell-
CC nerve cell interaction.
SQ Sequence 861 AA;

Query Match          71.3%; Score 57; DB 22; Length 861;
Best Local Similarity 75.0%; Pred. No. 7.20e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 270 kwtsflka 277
   :| |||||
Qy 248 KWNTEFLKA 255

RESULT 7
ID W58540 standard; Protein; 861 AA.
AC W58540;
DE 02-SEP-1998 (first entry)

```

```

DE Human semaphorin.
KW Human; semaphorin; diagnosis; nervous disease; immune disease;
KW nerve extension inhibiting activity.
OS Homo sapiens.
PN J10155490-A.
PD 16-JUN-1998.
PF 27-NOV-1996; 332900.
PR 27-NOV-1996; JP-332900.
PA (SUMU ) SUMITOMO SEIYAKU KK.
DR WPI; 98-391044/34.
DR N-PSDB; V31121.
PT New human semaforin gene - useful in the diagnosis of nervous system
PT and immune disorders
PS Claim 1; Page 10-12; 15pp; Japanese.
CC The present sequence represents human semaphorin (translated from the
CC Japanese specification as semaforin). Semaphorin has nerve extension
CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
CC treatment and researches on nervous diseases and immune diseases.
SQ Sequence 861 AA;

```

```

Query Match          71.3%; Score 57; DB 32; Length 861;
Best Local Similarity 75.0%; Pred. No. 7.20e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```
Db 270 kwtsflka 277
```

```
Qy 248 KWNTEFLKA 255
```

```

RESULT 8
ID W17657 standard; Protein; 862 AA.
AC W17657;
DE 24-JUL-1997 (first entry)
DE Human CD100 antigen.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
KW vaccine.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..41
FT /label= Sig_peptide
FT protein 42..862
FT /label= Mat_protein
FT domain 42..553
FT /label= Semaphorin_domain
FT domain 554..630
FT /label= Ig-like_domain
FT domain 631..733
FT /label= Stalk_domain
FT domain 735..752
FT /label= Transmembrane_domain
FT domain 753..862
FT /label= Cytoplasmic_domain
FT modified_site 808..815
FT /label= Phosphorylation
FT /note= "putative tyrosine phosphorylation site"
PN W09717368-AL.
PD 15-MAY-1997.
PR 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND ) DANA FARBER CANCER INST.
PI Boussioutis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI; 97-280982/25.
DR N-PSDB; T60665.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Claim 7; Page 70-72; 135pp; English.
CC Human CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60665) isolated by COS cell expression cloning. CD100
CC polypeptides and fusion proteins, nucleic acids, and host cells

```

CC expressing CD100 can be utilised in diagnostic and therapeutic
 CC methods involving modulation of B and T cell responses, neuron
 CC axonal growth and immune cell-nerve cell interaction.
 SQ Sequence 862 AA;

Query Match 71.3%; Score 57; DB 22; Length 862;
 Best Local Similarity 75.0%; Pred. No. 7.20e+01;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 270 kwtsflka 277

II:|||||
 QY 248 KWNTEFLKA 255

RESULT 9

ID R26167 standard; Protein; 340 AA.

AC R26167;

DT 02-FEB-1993 (first entry)

DE 85-C.

KW Immunise; antibody; vaccine; PCR; antisera; amplify.

OS Mycobacterium tuberculosis.

FH Key Location/Qualifiers

FT peptide 1..46

FT protein /label.. Signal_peptide

FT 47..340

FT /label.. Mature_peptide

PN EP-499003-A.

PD 19-AUG-1992.

PF 14-FEB-1991; 400388.

PR 14-FEB-1991; EP-400388.

PA (INNO-) INNOGENETICS NV SA.

PI Content J, De Bruyn J, De Wit L;

DR WPI: 92-277793/34.

DR N-PSDB: Q27378.

PT Recombinant peptide(s) and their nucleic acids - for diagnosing

PT tuberculosis and as a vaccine against tuberculosis

PS Disclosure; Fig 1; 48pp; English.

CC The sequence given is 85-C from Mycobacterium tuberculosis. The 85-C

gene was amplified from genomic DNA using the primer sequences given

CC in Q27373-4. The 85-C gene was sequenced and the amino acid sequence

CC determined. Various peptide motifs were identified and could then be

CC used in the production of vaccines for immunisation against

CC tuberculosis. The peptides may be used to raise antisera or antibodies

CC against tuberculosis.

SQ Sequence 340 AA;

Query Match 67.5%; Score 54; DB 5; Length 340;

Best Local Similarity 60.0%; Pred. No. 1.38e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 140 kwetfltrm 149

II:|||||

QY 248 KWNTEFL-KAM 256

RESULT 10

ID R71381 standard; Protein; 441 AA.

AC R71381;

DT 21-NOV-1995 (first entry)

DE Vaccinia virus semaphorin IV protein.

KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;

KW varicella major virus; smallpox; semaphorin receptor binding activity;

KW modulation; nerve cell growth; immune response; viral pathogenesis;

KW neurological disease; neuro-regeneration; oncological infection.

OS Vaccinia virus.

PN W09507706-A.

PD 23-MAR-1995.

PF 13-SEP-1994; U10151.

PR 13-SEP-1993; US-121713.

PA (REGC) UNIV CALIFORNIA.

PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;

PI O'Connor T;

DR WPI: 95-131177/17.

DR N-PSDB: Q87443.
 PT New class of semaphorin peptide(s) and polypeptide(s) - are
 PT potent modulators of nerve cell growth and regeneration
 PS Example 2; Page 65-67; 101pp; English.
 CC The sequence of the vaccinia virus semaphorin IV protein.
 CC The gene sequence was isolated as the A39R open reading frame sequence
 CC from variola, based on sequence homology searches of a database with the
 CC grasshopper, tribolium and drosophila semaphorin sequences. The proteins
 CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III
 CC (Q87442), vaccinia virus semaphorin I or varicella major (smallpox) virus
 CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
 CC virus semaphorin IV (Q87447) genes were used to generate a series of
 CC peptides (R70370-R70418), which retain semaphorin receptor binding
 CC activity. The semaphorin derived or semaphorin receptor derived peptides
 CC are potent modulators of nerve cell growth, immune responsiveness and
 CC viral pathogenesis. They can be used in diagnosis and treatment of
 CC neurological disease and neuro-regeneration, immune modulation and
 CC diagnosis and treatment of viral and oncological infection and diseases.
 SQ Sequence 441 AA;

Query Match 67.5%; Score 54; DB 13; Length 441;

Best Local Similarity 71.4%; Pred. No. 1.38e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 247 rwstflk 253

II:|||||

QY 248 KWNTEFLK 254

RESULT 11

ID R71384 standard; Protein; 712 AA.

AC R71384;

DT 21-NOV-1995 (first entry)

DE Tribolium semaphorin I protein.

KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;

KW varicella major virus; smallpox; semaphorin receptor binding activity;

KW modulation; nerve cell growth; immune response; viral pathogenesis;

KW neurological disease; neuro-regeneration; oncological infection.

OS Tribolium sp.

PN W09507706-A.

PD 23-MAR-1995.

PF 13-SEP-1994; U10151.

PR 13-SEP-1993; US-121713.

PA (REGC) UNIV CALIFORNIA.

PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;

PI O'Connor T;

DR WPI: 95-131177/17.

DR N-PSDB: Q87446.

PT New class of semaphorin peptide(s) and polypeptide(s) - are

PT potent modulators of nerve cell growth and regeneration

PS Example 2; Page 85-88; 101pp; English.

CC The sequence of the beetle Tribolium semaphorin I protein. The gene was

CC isolated by PCR using Tribolium genomic DNA. The proteins encoded by the

CC grasshopper semaphorin I (Q87441), human semaphorin III (Q87442),

CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II

CC (Q87444-5), Tribolium semaphorin I or varicella major (smallpox) virus

CC semaphorin IV (Q87447) genes were used to generate a series of peptides

CC (R70370-R70418), which retain semaphorin receptor binding activity. The

CC semaphorin derived or semaphorin receptor derived peptides are potent

CC modulators of nerve cell growth, immune responsiveness and viral

CC pathogenesis. They can be used in diagnosis and treatment of neurological

CC disease and neuro-regeneration, immune modulation and diagnosis and

CC treatment of viral and oncological infection and diseases.

SQ Sequence 712 AA;

Query Match 67.5%; Score 54; DB 13; Length 712;

Best Local Similarity 62.5%; Pred. No. 1.38e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 255 rwstflka 262

II:|||||

QY 248 KWNTEFLKA 255


```

RESULT 12
ID W92233 standard; Protein; 554 AA.
AC W92233;
DE 16-OCT-1998 (first entry)
DT HAP4 protein.
KW HAP4 gene; carbon metabolism pathway; metabolic pathway activation;
KW flavour-enhancer; metabolite production; yeast biomass enhancer.
OS Saccharomycetes cerevisiae.
PN W09826079-A1.
PD 18-JUN-1998.
PF 12-DEC-1997; NL0688.
PR 12-DEC-1996; EP-203520.
PA (UNAM ) UNIV AMSTERDAM.
PI Blom J, Grievell LA, Teixeira De Mattos MJ;
DR WPI: 98-377265/32.
DR N-PSDB: V44782, V44783, V44784, V44785.
PT Activation of glucose-repressed pathways in yeast in presence of
PT glucose - by introduction of activator protein into cell, useful for
PT increased biomass, e.g. for heterologous protein production
PS Disclosure; Fig 6; 72pp; English.
CC This sequence is the yeast HAP4 protein. It can be used in the
CC method of the invention for providing a microorganism which has a
CC preferred metabolic pathway in the presence of a certain carbon source,
CC with the ability to inhibit or circumvent that pathway, comprises
CC providing the microorganism with ability to derepress or circumvent the
CC repression of the non-preferred pathway in the presence of the C source.
CC The method allows the culturing of an increased biomass of yeast per
CC consumed unit of glucose. This has applications in the baking industry,
CC as sources of flavour-enhancing yeast extracts, for the production of
CC metabolites and heterologous gene products, e.g. enzymes, precursors for
CC chemicals, biosurfactants, fatty acids, especially for the
CC pharmaceutical, agricultural and food sectors. The method allows the
CC activation of metabolic pathways, e.g. the oxidative pathways, even in
CC the presence of glucose. It also results in a decreased amount of the
CC fermentative by-product ethanol.
CC Sequence 554 AA;
SQ

Query Match 66.3%; Score 53; DB 34; Length 554;
Best Local Similarity 50.0%; Pred. No. 1.71e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 336 wnsylqsm 343
QY 249 WNTFLKAM 256
|||:|:|

RESULT 13
ID R71382 standard; Protein; 650 AA.
AC R71382;
DE 21-NOV-1995 (first entry)
DE Drosophila semaphorin I protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Drosophila sp.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC ) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB: Q87444.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 74-77; 101pp; English.
CC The sequence of the Drosophila semaphorin I protein. The gene was
CC isolated by PCR using primers based on sequence homology between the
CC complete grasshopper and partial Tribolium and Manduca (moth) semaphorin
CC gene sequences. The products gave two different sequences, each of which

```

was used to obtain its respective complete sequence: semaphorin I and II (Q87445). The proteins encoded by the grasshopper semaphorin I (Q87441), human semaphorin III (Q87442), vaccinia virus semaphorin IV, Drosophila semaphorin I and II, Tribolium semaphorin I (Q87446) or variola major (smallpox) virus semaphorin IV (Q87447) genes were used to generate a series of peptides (R70370-R70418), which retain semaphorin receptor binding activity. The semaphorin derived or semaphorin receptor derived peptides are potent modulators of nerve cell growth, immune responsiveness and viral pathogenesis. They can be used in diagnosis and treatment of neurological disease and neuro-regeneration, immune modulation and diagnosis and treatment of viral and oncological infection and diseases.

Query Match 66.3%; Score 53; DB 13; Length 650;
Best Local Similarity 44.4%; Pred. No. 1.71e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 153 nrwtsfiks 161
QY 247 SKWNTFLKA 255
:|||:|:

RESULT 14
ID W75605 standard; peptide; 15 AA.
AC W75605;
DE 23-OCT-1998 (first entry)
DE M. tuberculosis 30 kD protein derived peptide 19 (residues 91-105).
KW Mycobacterium tuberculosis; vaccination; extracellular product;
KW immunodominant epitope; interleukin-12; MF59; immune response;
KW opsonising humoral response; intracellular pathogen.
OS Synthetic.
OS Mycobacterium tuberculosis.
PN W09831388-A1.
PD 23-JUL-1998.
PF 15-JAN-1998; U00942.
PR 21-JAN-1997; US-786533.
PA (REGC) UNIV CALIFORNIA.
PI Harth G, Horwitz MA, Lee B;
DR WPI: 98-413815/35.
PT Vaccines against Mycobacterium containing major extracellular
PT proteins - used to, e.g. induce protective and therapeutic immune
PT responses, and for detecting an immune response
PS Example 28; Page 96; 236pp; English.
CC Sequences shown in W75587 to W75641 represent synthetic peptides derived
CC from the native 30kD major secretory protein of M. tuberculosis. These
CC peptides are used for splenic lymphocyte proliferation assays to identify
CC the immunodominant T-cell epitope of the 30kD protein. The invention
CC provides an agent for vaccinating mammals against Mycobacterium. The
CC agent comprises at least one of the major abundant extracellular 110,
CC 80, 71, 58, 45, 32A, 32B, 30, 24, 23.5, 23, 16, 14 or 12 kD proteins of
CC M. tuberculosis, or at least 1 of their immunodominant epitopes and
CC interleukin-12 (IL-12) or MF59 as adjuvants. The agent containing the
CC nucleic acid encoding the extracellular products are used to raise a
CC protective or therapeutic immune response against Mycobacterium,
CC specifically M. tuberculosis. The immunodominant epitopes can also be
CC used (typically in a cutaneous hypersensitivity test) to detect an
CC immune response to vaccination. Preparation of the agent does not require
CC selection of the most immunogenic products, so large scale production and
CC purification are easy, resulting in a consistent, standardised
CC formulation, having lower toxicity than killed or attenuated vaccines.
CC The agents provide a rapid and effective response (including a strong
CC cell-mediated component) and are safe even in immunocompromised subjects.
CC They prevent development of an opsonising humoral response that might
CC spread intracellular pathogens.
SQ Sequence 15 AA;

Query Match 65.0%; Score 52; DB 34; Length 15;
Best Local Similarity 62.5%; Pred. No. 2.11e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 6 kwetflts 13
|||:|:

QY 248 KWNTFLKA 255

RESULT 15
ID W75660 standard; peptide; 15 AA.
AC W75660;
DT 23-OCT-1998 (first entry)
DE M. tuberculosis 32A kD protein derived peptide 19 (residues 91-105).
KW Mycobacterium tuberculosis; vaccination; extracellular product;
KW immunodominant epitope; interleukin-12; MF59; immune response;
KW opsonising humoral response; intracellular pathogen.
OS Synthetic.
OS Mycobacterium tuberculosis.
PN WO9831388-A1.
PD 23-JUL-1998.
PF 15-JAN-1998; U00942.
PR 21-JAN-1997; US-786533.
PA (REGC) UNIV CALIFORNIA.
PI Harth G, Horwitz MA, Lee B;
DR WPI; 98-413815/35.
PT Vaccines against Mycobacterium containing major extracellular
PT proteins - used to, e.g. induce protective and therapeutic immune
PT responses, and for detecting an immune response
PS Cidm 13; Page 203; 236pp; English.
CC This is an immunodominant peptide epitope from the M. tuberculosis 32A kD
CC protein. Sequences shown in W75642 to W75698 represent synthetic peptides
CC derived from the native 32A kD major secretory protein of Mycobacterium
CC tuberculosis. These peptides are used for identifying the immunodominant
CC T-cell epitope of the 32A kD protein. The invention provides an agent for
CC vaccinating mammals against Mycobacterium. The agent comprises at least
CC one of the major abundant extracellular 110, 80, 71, 58, 45, 32A, 32B,
CC 30, 24, 23.5, 23, 16, 14 or 12 kD proteins of M. tuberculosis, or at
CC least 1 of their immunodominant epitopes and interleukin-12 (IL-12) or
CC MF59 as adjuvants. The agent containing the nucleic acid encoding the
CC extracellular products are used to raise a protective or therapeutic
CC immune response against Mycobacterium, specifically M. tuberculosis.
CC The immunodominant epitopes can also be used (typically in a cutaneous
CC hypersensitivity test) to detect an immune response to vaccination.
CC Preparation of the agent does not require selection of the most
CC immunogenic products, so large scale production and purification are
CC easy, resulting in a consistent, standardised formulation, having lower
CC toxicity than killed or attenuated vaccines. The agents provide a rapid
CC and effective response (including a strong cell-mediated component) and
CC are safe even in immunocompromised subjects. They prevent development of
CC an opsonising humoral response that might spread intracellular pathogens.
SQ Sequence 15 AA;

Query Match 65.0%; Score 52; DB 34; Length 15;
Best Local Similarity 62.5%; Pred. No. 2.11e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 6 kwetflts 13

QY 248 KWNTFLKA 255

Search completed: Thu Jul 8 18:43:53 1999
Job time : 18 secs.

W P E R L H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:43:03 1999; MasPar time 5.33 Seconds
75.203 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (247-256) from US09041236.pep (18 of 45)
Perfect Score: 80
Sequence: 1 SKWNTFLKAM 10

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 25.167; Variance 38.267; scale 0.658

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	62	77.5	753	2	G02173 semaphorin III family	7.03e-01
2	61	76.3	666	2	I58169 semaphorin III - mouse	1.06e+00
3	61	76.3	748	2	I48744 semaphorin A - mouse	1.06e+00
4	61	76.3	749	2	G01856 semaphorin V - human	1.06e+00
5	61	76.3	751	2	I48748 semaphorin E - mouse	1.06e+00
6	61	76.3	771	2	D49423 semaphorin III precu	1.06e+00
7	61	76.3	772	2	I48747 semaphorin D - mouse	1.06e+00
8	61	76.3	772	2	A09069 collapsin - chicken	1.06e+00
9	60	75.0	653	2	T03102 semaphorin homolog A3	1.60e+00
10	60	75.0	760	2	I48745 semaphorin B - mouse	1.60e+00
11	60	75.0	834	2	S66498 M-sema F protein prec	1.60e+00
12	59	73.8	335	2	E71215 hypothetical protein	2.40e+00
13	57	71.3	4273	2	C69679 polyketide synthase p	5.35e+00
14	56	70.0	254	1	BVBV22 pRT122 protein precu	7.93e+00
15	56	70.0	354	2	S28755 hypothetical protein	7.93e+00
16	56	70.0	411	2	A34526 ORF1 protein - Orgyia	7.93e+00
17	56	70.0	511	2	S49151 maturase matk, intron	7.93e+00
18	56	70.0	542	2	J02021 hypothetical 65k prot	7.93e+00
19	56	70.0	544	2	S58532 matk protein (trnk in	7.93e+00
20	56	70.0	896	1	GNLJGH pol polyprotein - hum	7.93e+00
21	56	70.0	896	1	GNLJGH pol polyprotein - hum	7.93e+00
22	55	68.8	315	2	I39479 hypothetical protein	1.17e+01
23	55	68.8	337	2	E69852 malate dehydrogenase	1.17e+01

24	55	68.8	714	2	S68603 hypothetical protein	1.17e+01
25	55	68.8	797	2	D71621 hypothetical protein	1.17e+01
26	54	67.5	295	2	J01775 Sall9R protein - vacc	1.72e+01
27	54	67.5	330	2	JN0897 alpha-antigen precurs	1.72e+01
28	54	67.5	340	2	D70615 antigen fbpc2 - Mycob	1.72e+01
29	54	67.5	403	2	E42521 A39R protein - vaccin	1.72e+01
30	54	67.5	428	2	B69187 conserved hypotheticala	1.72e+01
31	54	67.5	437	2	S67305 phosphopyruvate hydra	1.72e+01
32	54	67.5	437	2	S69881 phosphopyruvate hydra	1.72e+01
33	54	67.5	441	2	S29921 hypothetical protein	1.72e+01
34	54	67.5	442	2	C69785 cellobiose phosphotra	1.72e+01
35	54	67.5	475	1	I24874 helicase (EC 3.6.1.-)	1.72e+01
36	54	67.5	711	2	A49423 semaphorin I precurs	1.72e+01
37	54	67.5	782	2	I48746 semaphorin C - mouse	1.72e+01
38	53	66.3	272	2	F71162 hypothetical protein	2.51e+01
39	53	66.3	382	1	J01291 matrix protein - para	2.51e+01
40	53	66.3	382	1	J01290 matrix protein - para	2.51e+01
41	53	66.3	434	2	S42676 KES1 protein - yeast	2.51e+01
42	53	66.3	434	2	S42677 HES1 protein - yeast	2.51e+01
43	53	66.3	554	2	S37936 regulatory protein HA	2.51e+01
44	53	66.3	564	2	B64213 methylgalactoside per	2.51e+01
45	53	66.3	1145	1	GNLJEV pol polyprotein - equ	2.51e+01

ALIGNMENTS

RESULT 1

ENTRY G02173 #type complete
TITLE semaphorin III family homolog - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1998 #sequence_revision 06-Jun-1997 #text_change 04-Sep-1998

ACCESSIONS G02173
REFERENCE G09275
#authors Naylor, S.
#submission submitted to the EMBL Data Library, October 1995
#accession G02173
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-753 ##label NAY
##cross-references EMBL:U38276; NID:g1061350; PID:g1061351
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 753 #molecular-weight 84941 #checksum 5681

Query Match 77.5%; Score 62; DB 2; Length 753;
Best Local Similarity 77.8%; Pred. No. 7.03e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 280 NWSTFLKA 288
QY 247 SKWNTFLKA 255
:::|||||

RESULT 2

ENTRY I58169 #type fragment
TITLE semaphorin III - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 04-Sep-1998

ACCESSIONS I58169
REFERENCE I58169
#authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.; Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#journal Neuron (1995) 14:949-959
#title Semaphorin III can function as a selective chemorepellent to pattern sensory projections in the spinal cord.
#cross-references MUID:95267432
#accession I58169
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-666 ##label RES
##cross-references GB:L40484; NID:g703189; PID:g703190
GENETICS

```
#gene Semaphorin V - mouse
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 666 #checksum 9654

Query Match 76.3%; Score 61; DB 2; Length 666;
Best Local Similarity 77.8%; Pred. No. 1.06e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 175 NKWTFELKA 183
QY 247 SKWTFELKA 255
:|:|:|:|:|:|

RESULT 3
ENTRY I48744 #type complete
TITLE semaphorin A - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48744
REFERENCE Puschel, A.W.; Adams, R.H.; Betz, H.
#authors Neuron (1995) 14:941-948
#journal Murine semaphorin D/collapsin is a member of a diverse gene
#title family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48744
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-748 #label RES
#cross-references EMBL:X85990; NID:g854323; PID:g854324
GENETICS
#gene sema
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 748 #molecular-weight 82894 #checksum 9017

Query Match 76.3%; Score 61; DB 2; Length 748;
Best Local Similarity 77.8%; Pred. No. 1.06e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 280 NKWTFELKA 288
QY 247 SKWTFELKA 255
:|:|:|:|:|:|

RESULT 4
ENTRY G01856 #type complete
TITLE semaphorin V - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
04-Sep-1998
ACCESSIONS G01856
REFERENCE Sekido, Y.
#authors Submitted to the EMBL Data Library, June 1995
#accession G01856
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-749 #label SEK
#cross-references EMBL:U28369; NID:g974283; PID:g974284
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 749 #molecular-weight 83121 #checksum 2747

Query Match 76.3%; Score 61; DB 2; Length 749;
Best Local Similarity 77.8%; Pred. No. 1.06e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 281 NKWTFELKA 289
QY 247 SKWTFELKA 255
:|:|:|:|:|:|

RESULT 5
ENTRY G01856 #type complete
TITLE semaphorin V - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
04-Sep-1998
ACCESSIONS G01856
REFERENCE Sekido, Y.
#authors Submitted to the EMBL Data Library, June 1995
#accession G01856
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-749 #label SEK
#cross-references EMBL:U28369; NID:g974283; PID:g974284
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 749 #molecular-weight 83121 #checksum 2747

Query Match 76.3%; Score 61; DB 2; Length 749;
Best Local Similarity 77.8%; Pred. No. 1.06e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 281 NKWTFELKA 289
QY 247 SKWTFELKA 255
:|:|:|:|:|:|

RESULT 6
ENTRY D49423 #type complete
TITLE semaphorin III precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
04-Sep-1998
ACCESSIONS D49423
REFERENCE Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
#authors Cell (1993) 75:1389-1399
#journal The Semaphorin genes encode a family of transmembrane and
#title secreted growth cone guidance molecules.
#accession D49423
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-771 #label KOL
#cross-references GB:L26081; NID:g799328; PID:g436560
GENETICS
#gene GDB:SEMA1
#cross-references GDB:283448
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 771 #molecular-weight 88889 #checksum 5249

Query Match 76.3%; Score 61; DB 2; Length 771;
Best Local Similarity 77.8%; Pred. No. 1.06e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 281 NKWTFELKA 289
QY 247 SKWTFELKA 255
:|:|:|:|:|:|

RESULT 7
ENTRY I48747 #type complete
TITLE semaphorin D - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48747
REFERENCE Puschel, A.W.; Adams, R.H.; Betz, H.
#authors Neuron (1995) 14:941-948
#journal
```

#title Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48747

##status preliminary; translated from GB/EMBL/DBBJ
##molecule_type mRNA
##residues 1-772 ##label RES
##cross-references EMBL:X85993; NID:g854329; PID:g854330

GENETICS

#gene semD

CLASSIFICATION #superfamily semaphorin
SUMMARY #length 772 #molecular-weight 88710 #checksum 1776

Query Match 76.3%; Score 61; DB 2; Length 772;
Best Local Similarity 77.8%; Pred. No. 1.06e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 281 NKWTFLLKA 289

::: || || || || ||

QY 247 SKWTFLLKA 255

RESULT 8

ENTRY #type complete
TITLE collapsin - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 04-Sep-1998

ACCESSIONS

REFERENCE A49069

#authors Luo, Y.; Raible, D.; Raper, J.A.

#journal Cell (1993) 75:217-227

#title Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones.

#accession A49069

##status preliminary; not compared with conceptual translation

##molecule_type mRNA

##residues 1-772 ##label LUO

##cross-references GB:U02528; NID:g410078; PID:g410079

CLASSIFICATION #superfamily semaphorin

SUMMARY #length 772 #molecular-weight 88867 #checksum 9712

Query Match 76.3%; Score 61; DB 2; Length 772;

Best Local Similarity 77.8%; Pred. No. 1.06e+00;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 281 NKWTFLLKA 289

::: || || || || ||

QY 247 SKWTFLLKA 255

RESULT 9

ENTRY #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999

ACCESSIONS

REFERENCE T03102

#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.

#journal J. Virol. (1997) 71:6517-6525

#title Primary structure of the alcelaphine herpesvirus 1 genome.

#accession T03102

##status preliminary; translated from GB/EMBL/DBBJ

##molecule_type DNA

##residues 1-653 ##label ENS

##cross-references EMBL:AF005370; NID:g2337967; PID:g2337970

SUMMARY

#length 653 #molecular-weight 73645 #checksum 5501

Query Match 75.0%; Score 60; DB 2; Length 653;

Best Local Similarity 87.5%; Pred. No. 1.60e+00;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 302 KWTFLKA 309

::: || || || || ||

QY 248 KWTFLKA 255

RESULT 10

ENTRY #type complete
TITLE semaphorin B - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 28-Feb-1997

ACCESSIONS

REFERENCE I48745

#authors Puschel, A.W.; Adams, R.H.; Betz, H.

#journal Neuron (1995) 14:941-948

#title Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.

#cross-references MUID:95267431

#accession I48745

##status preliminary; translated from GB/EMBL/DBBJ

##molecule_type mRNA

##residues 1-760 ##label RES

##cross-references EMBL:X85991; NID:g854325; PID:g854326

GENETICS

#gene semB

SUMMARY #length 760 #molecular-weight 83458 #checksum 2188

Query Match 75.0%; Score 60; DB 2; Length 760;

Best Local Similarity 87.5%; Pred. No. 1.60e+00;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 282 KWTFLKA 289

::: || || || || ||

QY 248 KWTFLKA 255

RESULT 11

ENTRY #type complete
TITLE M-sema F protein precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 10-Sep-1997

ACCESSIONS

REFERENCE S66498

#authors Inagaki, S.; Furuyama, T.; Iwahashi, Y.

#journal FEBS Lett. (1995) 370:269-272

#title Identification of a member of mouse semaphorin family.

#cross-references MUID:95385809

#accession S66498

##status preliminary

##molecule_type mRNA

##residues 1-834 ##label INA

##cross-references EMBL:S79463; NID:g110598; PID:g110599

FEATURE

1-21 #domain signal sequence #status predicted #label SIG\
22-834 #product M-sema F protein #status predicted #label MAT
SUMMARY #length 834 #molecular-weight 92556 #checksum 7189

Query Match 75.0%; Score 60; DB 2; Length 834;

Best Local Similarity 87.5%; Pred. No. 1.60e+00;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 274 KWTFLKA 281

::: || || || || ||

QY 248 KWTFLKA 255

RESULT 12

ENTRY #type complete
TITLE hypothetical protein PH1989 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998

```

ACCESSIONS
REFERENCE
#authors
Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. (1998) 5:53-76
Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references MUID:98344137
#accession
#status
preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-335 #label KAW
#cross-references GB:AF000007; NID:g3236134; PID:d1032059; PID:g3258433
#experimental_source strain OT3
#note this accession replaces an interim accession for a
sequence replaced by GenBank
GENETICS
#gene
PH1989
SUMMARY
#length 335 #molecular-weight 37502 #checksum 731
Query Match 73.8%; Score 59; DB 2; Length 335;
Best Local Similarity 75.0%; Pred. No. 2.40e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 55 RWNTFLAK 62
:|||||:
QY 248 KWNFTLKA 255
C69679 #type complete
polyketide synthase pksM - Bacillus subtilis
#formal_name Bacillus subtilis
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
12-Feb-1999
C69679
A69580
#authors
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujica,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glasner, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Pressac, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takanashi, H.;
Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.;

```

```

Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession C69679
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-4273 #label KUN
#cross-references GB:299113; GB:AL009126; NID:g2634090; PID:e1183378;
PID:g2634103
#experimental_source strain 168
GENETICS
#gene
pksM
CLASSIFICATION
#superfamily short-chain alcohol dehydrogenase homology;
3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl
carrier protein homology
FEATURE
293-363
414-829
3145-3320
3550-3942
SUMMARY
#length 4273 #molecular-weight 477459 #checksum 7246
Query Match 71.3%; Score 57; DB 2; Length 4273;
Best Local Similarity 50.0%; Pred. No. 5.35e+00;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Db 2377 TKWGGFIKDM 2386
:|||||:
QY 247 SKWNFTLKA 256
BVB22 #type complete
PET122 protein precursor - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YER153c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 30-Jun-1991 #sequence_revision 30-Sep-1992 #text_change
18-Sep-1998
ACCESSIONS B36328; S50656; S05724; A46737; S12289
REFERENCE A36328
#authors Ohmen, J.D.; Burke, K.A.; McEwen, J.E.
#journal Mol. Cell. Biol. (1990) 10:3027-3035
#title Divergent overlapping transcripts at the PET122 locus in
Saccharomyces cerevisiae.
#cross-references MUID:90258894
#accession B36328
#molecule_type DNA
#residues 1-254 #label OHM
#cross-references GB:X07558
REFERENCE S50428
#authors Dietrich, F.S.
#submission submitted to the EMBL Data Library, December 1994
#description The sequence of S. cerevisiae cosmids 8229, 9115, 9132, 9981,
and lambda clones 7990 and 6134.
#accession S50656
#molecule_type DNA
#residues 1-254 #label DIE
#cross-references EMBL:U18917; NID:g603377; PID:g603393; MIPS:YER153C
S05724
REFERENCE
#authors Ohmen, J.D.; Kloeckener-Gruissem, B.; McEwen, J.E.
#journal Nucleic Acids Res. (1988) 16:10783-10802
#title Molecular cloning and nucleotide sequence of the nuclear

```

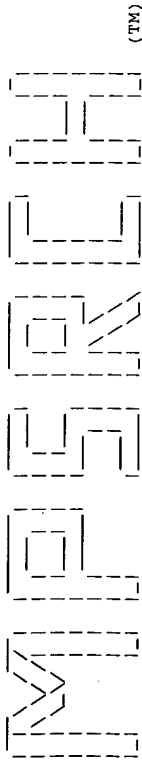
Search completed: Thu Jul 8 18:43:12 1999
Job time : 9 secs.

```

PEP122 gene required for expression of the mitochondrial
COX3 gene in S. cerevisiae.
#cross-references MUID:89083497
#accession S05724
##molecule_type DNA
##residues 1-198,'KT',201-222,'DVCWRSSVKVIA',235,'CRCWMPN' ##label
OH2
##cross-references EMBL:X07558
##note this sequence has been revised in reference A36328
REFERENCE A46737
#authors McMullin, T.W.; Fox, T.D.
#journal J. Biol. Chem. (1993) 268:11737-11741
#title COX3 mRNA-specific translational activator proteins are
associated with the inner mitochondrial membrane in
Saccharomyces cerevisiae.
#cross-references MUID:93280133
#accession A46737
##molecule_type protein
##residues 9-19 ##label MCM
GENETICS
#gene SGD:PET122
##cross-references SGD:S0000955; MIPS:YER153c
#map_position 5R
#genome nuclear
CLASSIFICATION #superfamily PET122 protein
KEYWORDS mitochondrion
FEATURE
1-8 #domain transit peptide (mitochondrion) #status
9-254 #product PET122 protein #status experimental #label MAT
SUMMARY #length 254 #molecular-weight 29155 #checksum 931
Query Match 70.0%; Score 56; DB 1; Length 254;
Best Local Similarity 55.6%; Pred.No. 7.93e+00;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 146 KWKFLOEM 154
QY 248 KWNTFLKAM 256

RESULT 15
ENTRY S28765 #type fragment
TITLE hypothetical protein (trnK intron) - barley chloroplast
(fragment)
ORGANISM #formal_name chloroplast Hordeum vulgare #common_name barley
DATE 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
09-Sep-1997
ACCESSIONS S28765
REFERENCE S28765
#authors Sexton, T.B.; Jones, J.T.; Mullet, J.E.
#journal Curr. Genet. (1990) 17:445-454
#title Sequence and transcriptional analysis of the barley ctDNA
region upstream of psbD-psbC encoding trnK(UUU), rps16,
trnQ(UUG), psbK, psbI, and trnS(GCU).
#accession S28765
##molecule_type DNA
##residues 1-354 #label SEX
##cross-references EMBL:X52765; NID:g11601; PID:g11602
GENETICS
#genome chloroplast
KEYWORDS chloroplast
SUMMARY #length 354 #checksum 3521
Query Match 70.0%; Score 56; DB 2; Length 354;
Best Local Similarity 50.0%; Pred.No. 7.93e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 189 SNMNSFITSM 198
QY 247 SKWNTFLKAM 256
```

THIS PAGE BLANK (USPTO)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:42:05 1999; MasPar time 3.30 Seconds
Tabular output not generated. 85.544 Million cell updates/sec

Title: >US-09-041-236-2
Description: (247-256) from US09041236.pep (18 of 45)
Perfect Score: 80
Sequence: 1 SKWNTFLKAM 10

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 25.946; Variance 34.540; scale 0.748

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	56	70.0	254	1	PT22_YEAST	2.78e+00
2	56	70.0	260	1	PT22_SACBA	2.78e+00
3	56	70.0	387	1	VP48_NPVAC	2.78e+00
4	56	70.0	411	1	VP48_NPVOP	2.78e+00
5	56	70.0	511	1	MATK_HORVU	2.78e+00
6	56	70.0	542	1	MATK_ORISA	2.78e+00
7	56	70.0	544	1	MATK_MAIZE	2.78e+00
8	56	70.0	896	1	POL_HTLIA	2.78e+00
9	56	70.0	896	1	POL_HTLIC	2.78e+00
10	55	68.8	315	1	VN35_ROTTL	4.28e+00
11	55	68.8	421	1	YDM4_SCHPO	4.28e+00
12	54	67.5	340	1	A85C_MYCTU	4.28e+00
13	54	67.5	437	1	VA39_VACCC	6.54e+00
14	54	67.5	403	1	ERRI_YEAST	6.54e+00
15	54	67.5	441	1	VA39_VACCV	6.54e+00
16	54	67.5	475	1	VG41_BPT4	6.54e+00
17	53	66.3	382	1	VNAT_P14HB	9.95e+00
18	53	66.3	382	1	VNAT_P14HA	9.95e+00
19	53	66.3	434	1	HESI_YEAST	9.95e+00
20	53	66.3	434	1	KESI_YEAST	9.95e+00
21	53	66.3	554	1	HAP4_YEAST	9.95e+00
22	53	66.3	564	1	MGLA_MYCGE	9.95e+00
23	53	66.3	1000	1	Y083_CAEEL	9.95e+00

24	53	66.3	1145	1	POL_EIAVY	POL POLYPROTEIN [CONTA	9.95e+00
25	53	66.3	1146	1	POL_EIASC	POL POLYPROTEIN [CONTA	9.95e+00
26	53	66.3	1146	1	POL_EIAV9	POL POLYPROTEIN [CONTA	9.95e+00
27	52	65.0	122	1	FMNB_DESV9	FMN-BINDING PROTEIN.	1.50e+01
28	52	65.0	288	1	CC2H_PLAFK	CELL DIVISION CONTROL	1.50e+01
29	52	65.0	325	1	A85B_MYCTU	ANTIGEN 85-B PRECURSOR	1.50e+01
30	52	65.0	325	1	A85B_MYCKA	ANTIGEN 85-B PRECURSOR	1.50e+01
31	52	65.0	327	1	A85B_MYCLE	ANTIGEN 85-B PRECURSOR	1.50e+01
32	52	65.0	330	1	A85A_MYCLE	ANTIGEN 85-A PRECURSOR	1.50e+01
33	52	65.0	330	1	A85B_MYCAV	ANTIGEN 85-B PRECURSOR	1.50e+01
34	52	65.0	338	1	A85A_MYCTU	ANTIGEN 85-A PRECURSOR	1.50e+01
35	52	65.0	572	1	HLV1_AERSA	HEMOLYSIN 1.	1.50e+01
36	52	65.0	740	1	PSAB_SYXEN	PHOTOSYSTEM I P700 CHL	1.50e+01
37	52	65.0	1076	1	MSH3_ARATH	DNA MISMATCH REPAIR PR	1.50e+01
38	52	65.0	2109	1	RRPL_VSVJO	RNA POLYMERASE BETA SU	1.50e+01
39	52	65.0	4427	1	PKSL_BACSU	POTATIVE POLYKETIDE SY	1.50e+01
40	51	63.8	245	1	YG24_YEAST	HYPOTHETICAL 28.6 KD P	2.26e+01
41	51	63.8	423	1	MTB1_BACAM	MODIFICATION METHYLASE	2.26e+01
42	51	63.8	1106	1	DPOD_BOVIN	DNA POLYMERASE DELTA C	2.26e+01
43	51	63.8	1107	1	DPOD_HUMAN	DNA POLYMERASE DELTA C	2.26e+01
44	51	63.8	1122	1	DPOL_ADE07	DNA POLYMERASE (EC 2.7	2.26e+01
45	51	63.8	2238	1	RRPL_BUNYW	RNA POLYMERASE (EC 2.7	2.26e+01

ALIGNMENTS

RESULT 1
ID PT22_YEAST STANDARD; PRT; 254 AA.
AC P10355;

DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE PET122 PROTEIN PRECURSOR.

OS PET122 OR YER153C.

GN SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;

OC SACCHAROMYCETACEAE; SACCHAROMYCES.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-AB320;

RX MEDLINE; 89083497.

RA OHMEN J.D., KLOECKENER-GRIJSSSEN B., MCEWEN J.E.;

RT "Molecular cloning and nucleotide sequence of the nuclear PET122 gene

RT required for expression of the mitochondrial COX3 gene in S.

RT cerevisiae.;

RL NUCLEIC ACIDS RES. 16:10783-10802(1988).

RN [2]

RP REVISIONS, SEQUENCE FROM N.A.

RX MEDLINE; 90258894.

RA OHMEN J.D., BURKE K.A., MCEWEN J.E.;

RT "Divergent overlapping transcripts at the PET122 locus in

RT Saccharomyces cerevisiae.;

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C / AB972;

RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,

RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,

RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNNICKE-SMITH S.,

RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,

RA MOSDALE D., NAKAHARA K., NAMATH A., NORGEN R., OEFNER P., OH C.,

RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGEN T., SMITH V.,

RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;

RA SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [4]

RP SEQUENCE OF 9-19, AND SUBCELLULAR LOCATION.

RX MEDLINE; 93280133.

RA MCMULLIN T.W., FOX T.D.;

RT "COX3 mRNA-specific transnational activator proteins are associated

RT with the inner mitochondrial membrane in Saccharomyces cerevisiae.;

RL J. BIOL. CHEM. 268:11737-11741(1993).

CC -!- FUNCTION: REQUIRED FOR EXPRESSION OF THE MITOCHONDRIAL GENE FOR

CC CYTOCHROME C OXIDASE SUBUNIT III (COX3). PET122 SEEMS TO WORKS BY

```
CC DIRECTLY INTERACTING WITH THE SMALL RIBOSOMAL SUBUNIT TO PROMOTE
CC TRANSLATION INITIATION ON THE COXIII MRNA.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL; INNER MEMBRANE ASSOCIATED.
CC -!- SIMILARITY: REGION OF 59% SIMILARITY TO THE CATALYTIC DOMAIN OF
CC E. COLI ALANYL TRNA-SYNTHASE.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07558; G4128; ALT_SEQ.
DR EMBL; U18917; G603393; -.
DR PIR; B36328; BVBY22.
DR SGD; L0001397; PET122.
KW TRANSLATION REGULATION; ACTIVATOR; TRANSIT PEPTIDE; MITOCHONDRION;
KW INNER MEMBRANE.
FT TRANSIT 1 8 MITOCHONDRION.
FT CHAIN 9 254 PET122 PROTEIN.
FT DOMAIN 185 254 REGION ESSENTIAL FOR PET122 FUNCTION.
SQ SEQUENCE 254 AA; 29155 MW; 598661f6 CRC32;
Query Match 70.0%; Score 56; DB 1; Length 254;
Best Local Similarity 55.6%; Pred. No. 2.78e+00;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 146 KWKFVLOQM 154
||: || |
Qy 248 KWNTFLKAM 256

RESULT 2
ID PT22_SACBA STANDARD; PRT; 260 AA.
AC O13374;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PET122 PROTEIN PRECURSOR.
OS SACCHAROMYCES BAYANUS (YEAST).
GN EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:NRRL Y-12624;
RA COSTANZO M.C., FOX T.D.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: REQUIRED FOR EXPRESSION OF THE MITOCHONDRIAL GENE FOR
CC CYTOCHROME C OXIDASE SUBUNIT III (COX3). PET122 SEEMS TO WORKS BY
CC DIRECTLY INTERACTING WITH THE SMALL RIBOSOMAL SUBUNIT TO PROMOTE
CC TRANSLATION INITIATION ON THE COXIII MRNA (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL; INNER MEMBRANE ASSOCIATED (BY
CC SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF026394; G2583061; -.
KW TRANSLATION REGULATION; ACTIVATOR; TRANSIT PEPTIDE; MITOCHONDRION;
KW INNER MEMBRANE.
FT TRANSIT 1 8 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 9 260 PET122 PROTEIN.
SQ SEQUENCE 260 AA; 29016 MW; 1EE5E620 CRC32;
Query Match 70.0%; Score 56; DB 1; Length 260;
```

```
Best Local Similarity 55.6%; Pred. No. 2.78e+00;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 150 KWKFVLOQM 158
||: || |
Qy 248 KWNTFLKAM 256

RESULT 3
ID VP48_NPVAC STANDARD; PRT; 387 AA.
AC Q00732;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE P48 PROTEIN (P45).
GN P48.
OS AUTOGAPHA CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS (ACMNPV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
CC NUCLEOPOLYHEDROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE; 943031173.
RA AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERBER M., POSSEE R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus.";
RL VIROLOGY 202:586-605(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HR3;
RA LU A., CRAIG A., CARSTENS E.B.;
RL SUBMITTED (JUN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 1-100 FROM N.A.
RC STRAIN=HR3;
RX MEDLINE; 92410596.
RA LU A., CARSTENS E.B.;
RT "Nucleotide sequence and transcriptional analysis of the p80 gene of
RT Autographa californica nuclear polyhedrosis virus: a homologue of the
RT Orsayia pseudotsugata nuclear polyhedrosis virus capsid-associated
RT gene.";
RL VIROLOGY 190:201-209(1992).
RN [4]
RP SEQUENCE 387 AA; 45313 MW; 3F143B05 CRC32;
SQ SEQUENCE 387 AA; 45313 MW; 3F143B05 CRC32;
Query Match 70.0%; Score 56; DB 1; Length 387;
Best Local Similarity 50.0%; Pred. No. 2.78e+00;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 161 WDFILRPM 168
||: ||: ||
Qy 249 WNTFLKAM 256

RESULT 4
ID VP48_NPVOP STANDARD; PRT; 411 AA.
AC P24651;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
```

DE P48 PROTEIN.
GN O88YIA PSEUDOTSUGATA MULTICAPSID POLYHEDROSIS VIRUS (OPMNPV).
OS VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
OC NUCLEOPOLYHEDROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90188300.
RA RUSSELL R.L.Q., ROHRMANN G.F.;
RT "The p6.5 gene region of a nuclear polyhedrosis virus of Orgyia
RT pseudotsugata: DNA sequence and transcriptional analysis of four late
RT genes.";
RL J. GEN. VIROL. 71:551-560(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 9032722.
RA MULLER R., PEARSON M., RUSSELL R.L.Q., ROHRMANN G.F.;
RT "A capsid-associated protein of the multicapsid nuclear polyhedrosis
RT virus of Orgyia pseudotsugata: genetic location, sequence,
RT transcriptional mapping, and immunocytochemical characterization.";
RL VIROLOGY 176:133-144(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97271300.
RA AHRENS C.H., RUSSELL R.R., FUNK C.J., EVANS J., HARWOOD S.,
RA ROHRMANN G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";
RL VIROLOGY 229:381-399(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13959; G222198; -;
DR EMBL; U75930; G1911350; -;
DR PIR; A34526; A34526.
KW LATE PROTEIN.
SQ SEQUENCE 411 AA; 47880 MW; F527EB76 CRC32;

Query Match 70.0%; Score 56; DB 1; Length 411;
Best Local Similarity 50.0%; Pred. No. 2.78e+00;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 165 WDIFLPM 172
QY 249 WNTFLKAM 256
I:|:|:|
|:|:|:|

RESULT 5
ID MATK_HORVU STANDARD; PRT; 511 AA.
AC P17158;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PROBABLE INTRON MATURASE.
GN MATK OR YCF14.
OS HORDEUM VULGARE (BARLEY).
OC CHLOROPLAST.
OC EUKARYOTA; VIRIDIPLANTAE; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; HORDEUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MORDEN C.W.;
RA SUBMITTED (DBC-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
RL [2]
RN SEQUENCE OF 1-354 FROM N.A.
RX MEDLINE; 90291518.

RA SEXTON T.B., JONES J.T., MULLET J.E.;
RT "Sequence and transcriptional analysis of the barley ctDNA region
RT upstream of psbD-psbC encoding trnK(UUU), rps16, trnQ(UUG), psbK,
RT psbI, and trnS(GCU).";
RL CURR. GENET. 17:445-454(1990).
RN [3]
RP SEQUENCE OF 355-504 FROM N.A.
RX STRAIN=CV, MOREX;
RX MEDLINE; 88335566.
RA BOYER S.K., MULLET J.E.;
RT "Sequence and transcript map of barley chloroplast psbA gene.";
RL NUCLEIC ACIDS RES. 16:8184-8184(1988).
CC -|- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS.
CC -|- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED
CC BY MITOCHONDRIAL INTRONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64129; G510119; -;
DR EMBL; X52765; G11602; -;
DR EMBL; X07942; G829299; -;
DR PIR; S01291; S01291.
DR MENDEL; 5072; HORVU; ycf14; 1.
KW CHLOROPLAST; RNA PROCESSING.
FT CONFLICT 109 154
FT CHLOROPLAST; IPFSLRELSCPKEIKPFQNLRSIHSPFLDKFLHLDY
FT LSHIE -> NIILATGILSERKNTKVSFTLYSFNISLF
FT RROIFAGLSITVR (IN REF. 2).
FT RIVYLDIIGINDLVNPLN -> PYLVFGYVRYQ (IN
FT REF. 3).
SQ SEQUENCE 511 AA; 61327 MW; 82C226B5 CRC32;

Query Match 70.0%; Score 56; DB 1; Length 511;
Best Local Similarity 50.0%; Pred. No. 2.78e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 189 SNWNSFTSM 198
QY 247 SKWNTFLKAM 256
I:|:|:|:|
|:|:|:|:|

RESULT 6
ID MATK_ORYSA STANDARD; PRT; 542 AA.
AC P12175;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PROBABLE INTRON MATURASE.
GN MATK OR YCF14.
OS ORYZA SATIVA (RICE).
OC CHLOROPLAST.
OC EUKARYOTA; VIRIDIPLANTAE; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; ORYZA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV, NIPPONBARE;
RA SUGIURA M.;
RL SUBMITTED (JUL-1989) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP COMPLETE GENOME.
RX MEDLINE; 89364698.
RA HIRATSUKA J., SHIMADA H., WHITTIER R., ISHIBASHI T., SAKAMOTO M.,
RA MORI M., KONDO C., HONJI Y., SUN C.-R., MENG B.-Y., LI Y.-Q.,
RA KANO A., NISHIZAWA Y., HIRAI A., SHINOZAKI K., SUGIURA M.;
RT "The complete sequence of the rice (Oryza sativa) chloroplast genome:

RT intermolecular recombination between distinct tRNA genes accounts for
RL a major plasmid DNA inversion during the evolution of the cereals.";
RL MOL. GEN. GENET. 217:185-194(1989).

CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS.

CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED
CC BY MITOCHONDRIAL INTRONS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL: X15901; G11959; -
DR PIR: JQ0201; JQ0201.
DR MENDEL: 5355; ORYSA:ycf14.1.
KW CHLOROPLAST; MRNA PROCESSING.
SQ SEQUENCE 542 AA; 65027 MW; 6800F595 CRC32;

Query Match 70.0%; Score 56; DB 1; Length 542;
Best Local Similarity 50.0%; Pred. No. 2.78e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 220 SNWNSFITSM 229
Y 247 SKWNTFLKAM 256
I:|:|:|:|

RESULT 7
ID MATK_MAIZE STANDARD; PRT; 544 AA.
AC P48190;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DE PROBABLE INTRON MATURASE.
GN MATK OR YCF14.
OS ZEA MAYS (MAIZE).
OC CHLOROPLAST.
CC EUKARYOTA: VIRIDIPHYTES; STREPTOPHYTES; EMERYOPHYTES; TRACHEOPHYTES;
CC EUPHYLLOPHYTES; SPERMATOPHYTES; MAGNOLIOPHYTES; LILIOPSIDA; POALES;
CC POACEAE; ZEA.
CC [1]
CC SEQUENCE FROM N.A.
RX MEDLINE; 95395841.
RA MAIER R.M., NECKERMAN K., IGLOI G.L., KOESSEL H.;
RT "Complete sequence of the maize chloroplast genome: gene content,
RT hotspots of divergence and fine tuning of genetic information by
RT transcript editing";
RL J. MOL. BIOL. 251:614-628(1995).

CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS.
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED
CC BY MITOCHONDRIAL INTRONS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL: X86563; G902202; -
DR MAIZEDB; 118213; -
DR MENDEL: 5109; ZEAMA:ycf14.1.
KW CHLOROPLAST; MRNA PROCESSING.
SQ SEQUENCE 544 AA; 64800 MW; 37F8307C CRC32;

Query Match 70.0%; Score 56; DB 1; Length 544;
Best Local Similarity 50.0%; Pred. No. 2.78e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 220 SNWNSFITSM 229
Y 247 SKWNTFLKAM 256
I:|:|:|:|

RESULT 8
ID POL_HTLIA STANDARD; PRT; 896 AA.
AC P03362;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DE POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE RIBONUCLEASE H (EC 3.1.26.4)].
GN POL.
OS HUMAN T-CELL LEUKEMIA VIRUS TYPE I (STRAIN ATK) (HTLV-I).

OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; BLV-HTLV RETROVIRUSES.
RN [1]
RX MEDLINE; 83221647.

RA SEIKI M., HATTORI S., HIRAYAMA Y., YOSHIDA M.;
RT "Human adult T-cell leukemia virus: complete nucleotide sequence of
RT the provirus genome integrated in leukemia cell DNA.";
RL PROC. NATL. ACAD. SCI. U.S.A. 80:3618-3622(1983).

RN [2]
RP SEQUENCE OF 69-185 FROM N.A.
RX MEDLINE; 89210803.

RA BANGHAM C.R.M., DAENKE S., PHILIPS R.E., CRUICKSHANK J.K.,
RA BELL J.I.;
RT "Enzymatic amplification of exogenous and endogenous retroviral
RT sequences from DNA of patients with tropical spastic paraparesis.";
RL EMBO J. 7:4179-4184(1988).

CC -!- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE
CC DETERMINED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL: J02029; G331148; -
DR EMBL: X14144; G930258; -
DR PIR: A03961; GNLJGH.
DR PFAM: PF00075; rnsah; 1.
DR PFAM: PF00078; rvt; 1.
DR PFAM: PF00552; integrase; 1.
DR PFAM: PF00865; rve; 1.
DR HSP; P04585; IKLM.
DR POLYPROTEIN; HYDROLASE; ENDONUCLEASE; NUCLEASE; TRANSFERASE;
KW RNA-DIRECTED DNA POLYMERASE.
SQ SEQUENCE 896 AA; 100141 MW; A5C74F7E CRC32;

Query Match 70.0%; Score 56; DB 1; Length 896;
Best Local Similarity 85.7%; Pred. No. 2.78e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 438 WNTFLKT 444
Y 249 WNTFLKA 255
I:|:|:|:|

RESULT 9
ID POL_HTLIC STANDARD; PRT; 896 AA.
AC P14078;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49)];
DE RIBONUCLEASE H (EC 3.1.26.4)].
GN POL.
OS HUMAN T-CELL LEUKEMIA VIRUS TYPE I (CARIBBEAN ISOLATE) (HTLV-I).
OC VIRUSES; RETROVIRIDAE; RETROVIRIDAE; BLV-HTLV RETROVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88274338.
RA MALIK K.T.A.; EVEN J.; KARPAS A.;
RT "Molecular cloning and complete nucleotide sequence of an adult T
cell leukaemia virus/human T cell leukaemia virus type I
(ATLV/HTLV-I) isolate of Caribbean origin: relationship to other
members of the ATLV/HTLV-I subgroup.";
RL J. GEN. VIROL. 69:1695-1710(1988).
CC -!- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE
CC DETERMINED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D13784; G221869; .
DR PIR: C28136; GNLCJN.
DR PFAM; PF00075; naseh; 1.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF00552; integrase; 1.
DR PFAM; PF00665; rve; 1.
DR HSPF; P04585; 1KLM.
KW POLYPROTEIN; HYDROLASE; ENDONUCLEASE; NUCLEASE; TRANSFERASE;
KW RNA-DIRECTED DNA POLYMERASE.
SQ SEQUENCE 896 AA; 100199 MW; D6BCDD2E CRC32;

Query Match 70.0%; Score 56; DB 1; Length 896;
Best Local Similarity 85.7%; Pred. No. 2.78e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 438 WNTFLKT 444
|||||
Qy 249 WNTFLKA 255

RESULT 10
ID VN35_ROT11 STANDARD; PRT; 315 AA.
AC Q03244;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE NONSTRUCTURAL RNA-BINDING PROTEIN 35 (NS35) (NCVP3).
GN S7.
OS TURKEY ROTAVIRUS (SEROTYPE 7 / STRAIN TV-1).
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93134787.
RA PATTON J.T.; SALTER-CID L.; KALBACH A.; MANSELL E.A.; KATTOURA M.;
RT "Nucleotide and amino acid sequence analysis of the rotavirus
nonstructural RNA-binding protein NS35.";
RL VIROLOGY 192:438-446(1993).
CC -!- FUNCTION: ESSENTIAL FOR GENOME REPLICATION AND FOR THE FORMATION
CC OF THE VIROPLASM. IT MAY ALSO BE IMPORTANT IN VIRAL RNA PACKAGING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; L04533; G294724; .
KW NONSTRUCTURAL PROTEIN; RNA-BINDING.
FT DOMAIN 204 240 RNA-BINDING (POTENTIAL).
SQ SEQUENCE 315 AA; 35930 MW; 2D036527 CRC32;

Query Match 68.8%; Score 55; DB 1; Length 315;
Best Local Similarity 40.0%; Pred. No. 4.28e+00;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 266 TNWAFVKSM 275
:::|::|
Qy 247 SKWNTFLKAM 256

RESULT 11
ID YDWA_SCHPO STANDARD; PRT; 421 AA.
AC O13911;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 48.5 KD PROTEIN C23C11.04C IN CHROMOSOME I.
GN SPAC23C11.04C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA BROWN D.; CHURCHER C.M.; BARRELL B.G.; RAJANDREAM M.A.; WOOD V.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: TO C.ELEGANS F21D5.5, YEAST YMR156C AND ACMPV ORF33.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z98559; E334132; .
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 421 AA; 48477 MW; AA47E3DE CRC32;

Query Match 68.8%; Score 55; DB 1; Length 421;
Best Local Similarity 50.0%; Pred. No. 4.28e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 167 GWN5FLKDV 176
:::|::|
Qy 247 SKWNTFLKAM 256

RESULT 12
ID A85C_MYCTU STANDARD; PRT; 340 AA.
AC P31953;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ANTIGEN 85-C PRECURSOR (85C) (ANTIGEN 85 COMPLEX C) (FIBRONECTIN-
DE BINDING PROTEIN C).
GN FBPC.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETES; CORYNEBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ERDMANN;
RX MEDLINE; 91348869.
RA CONTENT J.; LA CUVELLERIE A.; DE WIT L.; VINCENT-LEVY-FREBAULT V.;
RA OOMS J.; DE BRUN J.;
RT "The genes coding for the antigen 85 complexes of Mycobacterium

RT tuberculosis and Mycobacterium bovis BCG are members of a gene
RT family: cloning, sequence determination, and genomic organization of
RL the gene coding for antigen 85-C of *M. tuberculosis*.";
RL INFECT. IMMUN. 59:3705-3212(1991).
CC -1- FUNCTION: PROTEINS OF THE ANTIGEN 85 COMPLEX ARE RESPONSIBLE
CC FOR THE HIGH AFFINITY OF MYCOBACTERIA TO FIBRONECTIN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: TO OTHER MYCOBACTERIUM A85 ANTIGENS AND TO THE
CC N-TERMINAL OF C. GLUTAMICUM P51 PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X57229; G48828; -.
CC PIR: A43603; A43603.
CC PIR: S15504; S15504.
CC PFAM: PF00756; Esterase; 1.
CC KW ANTIGEN; SIGNAL.
CC FT SIGNAL 1 46 POTENTIAL.
CC FT CHAIN 47 340 ANTIGEN 85-C.
CC SQ SEQUENCE 340 AA; 36757 MW; 5B42E846 CRC32;

Query Match 67.5%; Score 54; DB 1; Length 340;

Best Local Similarity 60.0%; Pred. No. 6.54e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 140 KWTFLETRM 149

||:|||||

QY 248 KWTFLE-KAM 256

RESULT 13
ID VA39_VACCC STANDARD; PRT; 403 AA.
AC P21062;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DE 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE PROTEIN A39.
GN A39R.
OS VACCINIA VIRUS (STRAIN COPENHAGEN).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91021027.
RA GOBEL S. J., JOHNSON G. P., PERKUS M. E., DAVIS S. W., WINSLOW J. P.,
RA PAOLETTI E.;
RT "the complete DNA sequence of vaccinia virus.";
RL VIROLOGY 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA GOBEL S. J., JOHNSON G. P., PERKUS M. E., DAVIS S. W., WINSLOW J. P.,
RA PAOLETTI E.;
RL VIROLOGY 179:517-563(1990).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M35027; G335517; -.
CC PIR: E42521; E42521.
CC SQ SEQUENCE 403 AA; 45741 MW; FE0B99AB CRC32;

Query Match 67.5%; Score 54; DB 1; Length 403;

Best Local Similarity 71.4%; Pred. No. 6.54e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 209 RWSTFLK 215

||:|||||

QY 248 KWTFLEK 254

RESULT 14

ID ERR1_YEAST STANDARD; PRT; 437 AA.

AC P42222;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE ENOLASE RELATED PROTEIN 1 (EC 4.2.1.11).

GN ERRI OR YMR323W OR YMR924.15.

OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;

OC SACCHAROMYCETACEAE; SACCHAROMYCES.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RA CHURCHER C.N., LOUIS E.J., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;

RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE OF 120-437 FROM N.A.

RC STRAIN=S288C / YPI;

RX MEDLINE; 95304851.

RA PRYDE F.E., HUCKLE T.C., LOUIS E.J.;

RT "sequence analysis of the right end of chromosome XV in *Saccharomyces*

RT *cerevisiae*: an insight into the structural and functional

RT significance of sub-telomeric repeat sequences.";

RL YEAST 11:371-382(1995).

CC -1- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE -> PHOSPHOENOLPYRUVATE

CC + H(2)O.

CC -1- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Z54141; G1072413; -.

DR EMBL: U23472; G775203; -.

DR SGD; L0002735; ERRI.

DR PROSITE; PS00164; ENOLASE; 1.

DR PFAM; PF00113; enolase; 1.

DR HSSP; P00924; INEL.

KW LYASE; GLYCOLYSIS; MAGNESIUM.

FT ACT_SITE 160 160 BY SIMILARITY.

FT METAL 247 247 MAGNESIUM (BY SIMILARITY).

FT METAL 296 296 MAGNESIUM (BY SIMILARITY).

FT METAL 321 321 MAGNESIUM (BY SIMILARITY).

SQ SEQUENCE 437 AA; 47312 MW; FE6E4B47 CRC32;

Query Match 67.5%; Score 54; DB 1; Length 437;

Best Local Similarity 50.0%; Pred. No. 6.54e+00;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 305 SSWSAFLKTU 314

||:|||||

QY 247 SKWTFLEK 256

RESULT 15

ID VA39_VACCV STANDARD; PRT; 441 AA.

AC P24764;

DT 01-MAR-1992 (REL. 21, CREATED)

DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)

DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)

```

DE PROTEIN A39.
GN A39R OR (SALL9R AND SALLF1R).
OS VACCINIA VIRUS (STRAIN WR).
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91310644.
RA AMEGADZIE B.Y., AHN B.-Y., MOSS B.;
RT "Identification, sequence, and expression of the gene encoding a Mr
RT 35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase.";
RL J. BIOL. CHEM. 266:13712-13718(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91259063.
RA SMITH G.L., CHAN Y.S., HOWARD S.T.;
RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RT the right inverted terminal repeat.";
RL J. GEN. VIROL. 72:1349-1376(1991).
CC -!- CAUTION: IN REF.2 THIS SEQUENCE IS IN TWO PART (SALL9R AND
CC SALLF1R) DUE TO THE PRESENCE OF FRAMESHIFT COMPARED TO THAT OF
CC REF.1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61187; G335798;
DR EMBL: D11079; G222726; ALT_SEQ.
DR EMBL: X57318; G62234;
DR PIR: S29921; S29921.
SQ SEQUENCE 441 AA; 50185 MW; 2C823A68 CRC32;

Query Match 67.5%; Score 54; DB 1; Length 441;
Best Local Similarity 71.4%; Pred. No. 6.54e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 247 RWSTFLK 253
QY 248 KWTFLK 254

Search completed: Thu Jul 8 18:42:12 1999
Job time : 7 secs.

```

THIS PAGE BLANK (USPTO)

M P R E L H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:42:32 1999; MasPar time 8.54 Seconds
63.945 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (247-256) from US09041236.pep (18 of 45)
Perfect Score: 80
Sequence: 1 SKWNTFLKAM 10

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprenbl9
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 25.863; Variance 39.577; scale 0.653

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
1	80	100.0	393 11 088371 SEMAPHORIN L (FRAGMENT)	6.67e-04
2	80	100.0	665 4 075326 SEMAPHORIN L	6.67e-04
3	65	81.3	528 5 076726 Y40D12A.1 PROTEIN.	4.89e-01
4	62	77.5	284 11 054948 SEMAPHORIN IV HOMOLOG	1.68e+00
5	62	77.5	753 4 013372 SEMAPHORIN III FAMILY	1.68e+00
6	62	77.5	754 11 088633 SEMAPHORIN IV ISOFORM	1.68e+00
7	62	77.5	775 11 P70275 SEMAPHORIN H.	1.68e+00
8	62	77.5	775 4 015041 KIAA0331.	1.68e+00
9	62	77.5	785 4 Q15704 SEMAPHORIN.	1.68e+00
10	62	77.5	785 11 088632 SEMAPHORIN IV ISOFORM	1.68e+00
11	62	77.5	785 4 Q13275 SEMAPHORIN IV.	1.68e+00
12	61	76.3	294 13 090664 COLLAPSID-3 (FRAGMENT)	2.52e+00
13	61	76.3	748 11 062177 SEMAPHORIN A PRECURSOR	2.52e+00
14	61	76.3	749 4 Q13214 SEMAPHORIN V.	2.52e+00
15	61	76.3	750 4 Q93018 SEMAPHORIN V.	2.52e+00
16	61	76.3	751 11 Q62181 SEMAPHORIN E PRECURSOR	2.52e+00
17	61	76.3	751 4 Q99985 SEMAPHORIN E.	2.52e+00
18	61	76.3	751 13 Q42236 COLLAPSID 3.	2.52e+00
19	61	76.3	761 13 Q90663 COLLAPSID-2.	2.52e+00
20	61	76.3	771 4 Q14593 SEMAPHORIN-III.	2.52e+00

21	61	76.3	772 11	062180 SEMAPHORIN D PRECURSOR	2.52e+00
22	61	76.3	772 13	Q90607 COLLAPSID.	2.52e+00
23	61	76.3	772 11	Q90607 SEMAPHORIN III/COLLAPSID	2.52e+00
24	60	75.0	299 13	Q90666 COLLAPSID-5 (FRAGMENT)	3.76e+00
25	60	75.0	653 14	Q64906 SIMILAR TO GENBANK ACC	3.76e+00
26	60	75.0	760 11	Q62178 SEMAPHORIN B PRECURSOR	3.76e+00
27	60	75.0	785 13	Q42237 COLLAPSID 5	3.76e+00
28	60	75.0	834 11	Q64151 SEMAPHORIN I (M-SEMA F	3.76e+00
29	59	73.8	335 1	057713 335AA LONG HYPOTHETICA	5.60e+00
30	58	72.5	829 5	Q44737 F57B10.6 PROTEIN.	8.29e+00
31	57	71.3	295 13	Q90665 COLLAPSID-4 (FRAGMENT)	1.22e+01
32	57	71.3	861 11	Q09126 SEMAPHORIN J (SEMAPHOR	1.22e+01
33	57	71.3	862 4	Q92854 SEMAPHORIN.	1.22e+01
34	57	71.3	4273 2	Q31781 POLYKETIDE SYNTHASE.	1.22e+01
35	56	70.0	387 14	Q92463 ACNVPV ORF103.	1.79e+01
36	56	70.0	731 14	Q82324 T-CELL LYMPHOTROPIC VI	1.79e+01
37	56	70.0	896 14	Q90043 POL.	1.79e+01
38	56	70.0	896 14	Q56622 POLYMERASE (FRAGMENT).	1.79e+01
39	56	70.0	896 14	Q82322 PUTATIVE FRAMESHIFTING	1.79e+01
40	56	70.0	1461 14	Q56228 PR GAG-PRO-POL.	1.79e+01
41	55	68.8	315 2	Q43888 PLASMOD PIP501 COPR, R	2.62e+01
42	55	68.8	315 14	Q55586 NSP2.	2.62e+01
43	55	68.8	337 2	Q34736 MALATE DEHYDROGENASE-L	2.62e+01
44	55	68.8	714 2	Q68005 TYPE I TOPOISOMERASE.	2.62e+01
45	55	68.8	1236 5	Q77346 MAL3P4.10 PROTEIN.	2.62e+01

ALIGNMENTS

RESULT 1	
ID 088371	PRELIMINARY; PRT; 393 AA.
AC 088371	
DT 01-NOV-1998	(TREMBLREL. 08, CREATED)
DT 01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN L (FRAGMENT).	
GN SEMAL.	
OS MUS MUSCULUS (MOUSE).	
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;	
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE; 98389619.	
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;	
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA	
RL viruses."	
RT GENOMICS 51:340-350(1998).	
DR EMBL; AF030699; G3523117; .	
FT NON_TER 393	
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;	

Query Match 100.0%; Score 80; DB 11; Length 393;
Best Local Similarity 100.0%; Pred. No. 6.67e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 277 SKWNTFLKAM 286
IIIIIIIIII

QY 247 SKWNTFLKAM 256
IIIIIIIIII

RESULT 2	
ID 075326	PRELIMINARY; PRT; 666 AA.
AC 075326	
DT 01-NOV-1998	(TREMBLREL. 08, CREATED)
DT 01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN L.	
GN SEMAL.	
OS HOMO SAPIENS (HUMAN).	
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;	
OC CATARRHINI; HOMINIDAE; HOMO.	
RN [1]	
RP SEQUENCE FROM N.A.	

```

RX MEDLINE; 98389619.
RA LANE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSEER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses.";
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030698; G3523115; -.
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 80; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 6.67e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 279 SKWNTFLKAM 288
QY 247 SKWNTFLKAM 256
:|||||

RESULT 3
ID O76726 PRELIMINARY; PRT; 528 AA.
AC O76726;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE Y40D12A.1 PROTEIN.
GN Y40D12A.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA SMITH A., HARMON G.;
RT "The sequence of C. elegans cosmid Y40D12A.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF078792; G3329640; -.
SQ SEQUENCE 528 AA; 61523 MW; 0E80870A CRC32;

Query Match 81.3%; Score 65; DB 5; Length 528;
Best Local Similarity 66.7%; Pred. No. 4.89e-01;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 350 NNSNFKTM 358
QY 248 KNTFLKAM 256
:|||||

RESULT 4
ID O54948 PRELIMINARY; PRT; 284 AA.
AC O54948;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)

```

```

DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV HOMOLOG (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA HU L.-J., DRABKIN H.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF036162; G2687638; -.
FT NONTER 284 284
SQ SEQUENCE 284 AA; 32151 MW; 8D9DBF11 CRC32;

Query Match 77.5%; Score 62; DB 11; Length 284;
Best Local Similarity 77.8%; Pred. No. 1.68e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 213 NKWSTFLKA 221
QY 247 SKWNTFLKA 255
:|||||

RESULT 5
ID Q13372 PRELIMINARY; PRT; 753 AA.
AC Q13372;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN III FAMILY HOMOLOG.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 96230324.
RA XIANG R.H., HENSEL C.H., GARCIA D.K., CARLSON H.C., KOK K., DALY M.C.,
RA KERBACHER K., DEN BERG A., VELDHIJS P., BUYS C.H., NAYLOR S.L.;
RT "Isolation of the human semaphorin III/F gene (SEMA3F) at chromosome
RT 3p21, a region deleted in lung cancer.";
RL GENOMICS 32:39-48(1996).
DR EMBL; U38276; G1061351; -.
DR PFAM; PF00047; I9; 1.
SQ SEQUENCE 753 AA; 84941 MW; BECBEBB0 CRC32;

Query Match 77.5%; Score 62; DB 4; Length 753;
Best Local Similarity 77.8%; Pred. No. 1.68e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 280 NKWSTFLKA 288
QY 247 SKWNTFLKA 255
:|||||

RESULT 6
ID O88633 PRELIMINARY; PRT; 754 AA.
AC O88633;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN IV ISOFORM A.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ECKHARDT F., MEYERHANS A.;
RT "Molecular cloning and expression pattern of a murine semaphorin
RT homologous to H-sema IV.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF080091; G3377768; -.
SQ SEQUENCE 754 AA; 85133 MW; EC05BDF3 CRC32;

```

```
Query Match 77.5%; Score 62; DB 11; Length 754;
Best Local Similarity 77.8%; Pred. No. 1.68e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 281 NKWSTFLKA 289
Qy 247 SKWNTFLKA 255
      :||:|||||

RESULT 7
ID P70275 PRELIMINARY; PRT; 775 AA.
AC P70275; O09078; O09079;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN H.
GN SEMA OR SEMA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C FC3H;
RX MEDLINE; 98175564.
RA CHRISTENSEN C.C., TARABYKINA S.S., HULGAARD E.E., KRAMEROV D.D.,
RA LUKANIDIN E.E.;
RA "Transcription of a novel mouse semaphorin gene, M-semaH, correlates
RT with the metastatic ability of mouse tumor cell lines.";
RL CANCER RES. 58:1238-1244(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C FC3H;
RA CHRISTENSEN C., KLINGELHOFFER J., TARABYKINA S., HULGAARD E.,
RA KRAMEROV D., LUKANIDIN E.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Z93947; E1169841; ALT_SEQ.
DR EMBL; Z80941; E1293507; -.
DR EMBL; Z93948; E311738; ALT_SEQ.
DR MGD; MGI:1097708; SEMA.
DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 775 AA; 89503 MW; A3BC9ADE CRC32;

Query Match 77.5%; Score 62; DB 11; Length 775;
Best Local Similarity 77.8%; Pred. No. 1.68e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 282 NKWSTFLKA 290
Qy 247 SKWNTFLKA 255
      :||:|||||

RESULT 8
ID O15041 PRELIMINARY; PRT; 775 AA.
AC O15041;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE KIAA0331.
GN KIAA0331.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 97349984.
RA NAGASE T., ISHRAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,
RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RL code for large proteins in vitro.";
RL DNA RES. 4:141-150(1997).
DR EMBL; AB002329; D1021627; -.

DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 775 AA; 89227 MW; 68DEF6FC CRC32;

Query Match 77.5%; Score 62; DB 4; Length 775;
Best Local Similarity 77.8%; Pred. No. 1.68e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 282 NKWSTFLKA 290
Qy 247 SKWNTFLKA 255
      :||:|||||

RESULT 9
ID Q15704 PRELIMINARY; PRT; 785 AA.
AC Q15704;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 96226360.
RA ROCHE J., BOLDOG F., ROBINSON M., ROBINSON L., VARELLA-GARCIA L.,
RA SWANTON M., WAGGONER B., FISHEL R., FRANKLIN W., GEMMILL R.,
RA DRABKIN H.;
RT "Distinct 3p21.3 deletions in lung cancer and identification of a new
RL human semaphorin.";
RL ONCOGENE 12:1289-1297(1996).
DR EMBL; U33920; GI000207; -.
DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 785 AA; 9971F14E CRC32;

Query Match 77.5%; Score 62; DB 4; Length 785;
Best Local Similarity 77.8%; Pred. No. 1.68e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 312 NKWSTFLKA 320
Qy 247 SKWNTFLKA 255
      :||:|||||

RESULT 10
ID O88632 PRELIMINARY; PRT; 785 AA.
AC O88632;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV ISOFORM B.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ECKHARDT F., MEYERHANS A.;
RT "Molecular cloning and expression pattern of a murine semaphorin
RT homologous to H-sema IV.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF080090; G3377766; -.
SQ SEQUENCE 785 AA; 88493 MW; 6FFCD8F5 CRC32;

Query Match 77.5%; Score 62; DB 11; Length 785;
Best Local Similarity 77.8%; Pred. No. 1.68e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 312 NKWSTFLKA 320
Qy 247 SKWNTFLKA 255
      :||:|||||
```

```
RESULT 11
ID Q3275 PRELIMINARY; PRT; 785 AA.
AC Q3275; Q13274;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DE SEMAPHORIN A PRECURSOR (SEM A).
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA NELSON J., BIERWALD T.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 394-436 FROM N.A.
RX TISSUE-PLACENTA;
RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT cancer deletion region and demonstrate distinct expression
RT patterns.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR EMBL; AC000063; G1569373; -.
DR EMBL; U32172; G995788; -.
DR EMBL; U32171; G995786; -.
DR PFAM; PF00047; ig; 1.
SQ SEQUENCE 785 AA; 88381 MW; CD175765 CRC32;

Query Match 77.5%; Score 62; DB 4; Length 785;
Best Local Similarity 77.8%; Pred. No. 1.69e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 312 NKWSTFLKA 320
QY 247 SKWTFELKA 255
:|||||

RESULT 12
ID Q90664 PRELIMINARY; PRT; 294 AA.
AC Q90664;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE COLLAPIN-3 (FRAGMENT).
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95329269.
RA LUO Y., SHEPHERD I., LI J., RENZI M.J., CHANG S., RAPER J.A.;
RT "A family of molecules related to collapsin in the embryonic chick
RT nervous system.";
RL NEURON 14:1131-1140(1995).
DR EMBL; U28241; G886811; -.
FT NON_TER 1 1
FT NON_TER 294 294
SQ SEQUENCE 294 AA; 33046 MW; DE564C1E CRC32;

Query Match 76.3%; Score 61; DB 13; Length 294;
Best Local Similarity 77.8%; Pred. No. 2.52e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 42 NKWTFELKA 50
QY 247 SKWTFELKA 255
:|||||

RESULT 13
ID Q62177 PRELIMINARY; PRT; 748 AA.
AC Q62177;
```

```
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN A PRECURSOR (SEM A).
GN SEMA OR SEMA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
RT and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC LOW LEVELS FOUND BETWEEN DAYS 10-12.
CC EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL
CC BIRTH.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85990; G854324; -.
DR MGD; MGI:107561; SEMA.
DR PFAM; PF00047; ig; 1.
KW SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KW DEVELOPMENTAL PROTEIN.
FT SIGNAL 1 26
FT CHAIN 27 748
FT DOMAIN 586 649
FT DOMAIN IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 748 AA; 82894 MW; A7E53A8D CRC32;

Query Match 76.3%; Score 61; DB 11; Length 748;
Best Local Similarity 77.8%; Pred. No. 2.52e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 280 NKWTFELKA 288
QY 247 SKWTFELKA 255
:|||||

RESULT 14
ID Q13214 PRELIMINARY; PRT; 749 AA.
AC Q13214;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96210603.
RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT cancer deletion region and demonstrate distinct expression
RT patterns.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR EMBL; U28369; G974284; -.
DR PFAM; PF00047; ig; 1.
SQ SEQUENCE 749 AA; 83121 MW; 6138C5D5 CRC32;

Query Match 76.3%; Score 61; DB 4; Length 749;
Best Local Similarity 77.8%; Pred. No. 2.52e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 281 NKWTFELKA 289
QY 247 SKWTFELKA 255
:|||||
```

QY 247 SKWNTFLKA 255

RESULT 15
ID Q93018 PRELIMINARY; PRT; 750 AA.
AC Q93018;
DT 01-FEB-1997 (TREMREL. 02, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA DANTE M., WAMSLEY P.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U73167; G2880035; -.
DR PFAM; PF00047; 19; 1.
SQ SEQUENCE 750 AA; 83034 MW; CBC87E34 CRC32;

Query Match 76.3%; Score 61; DB 4; Length 750;
Best Local Similarity 77.8%; Pred. No. 2.52e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 282 NKWTFLLKA 290
:|||||
QY 247 SKWNTFLKA 255

Search completed: Thu Jul 8 18:42:45 1999
Job time : 13 secs.

THIS PAGE BLANK (USPTO)

W P S R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPPerch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:45:49 1999; MasPar time 10.00 Seconds
Tabular output not generated. 21.266 Million cell updates/sec

Title: >US-09-041-236-2
Description: (282-291) from US09041236.pep (19 of 45)
Perfect Score: 87
Sequence: 1 GOWRTRVYG 10

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Watch 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 18.948; Variance 63.105; scale 0.300

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	66	75.9	393	30	Murine Ena-Vasp like	9.22e+00
2	66	75.9	541	30	Mammalian Ena (Mena)	9.22e+00
3	66	75.9	740	30	Murine Mena, Ena-Vasp	9.22e+00
4	66	75.9	783	30	Mouse neural Mena+ pr	9.22e+00
5	66	75.9	787	30	Mouse neural Mena+ p	9.22e+00
6	66	75.9	802	30	Mouse neural Mena+++	9.22e+00
7	57	65.5	875	1	C-terminal prepro-C-t	7.21e+01
8	57	65.5	875	13	Peptidyl C-terminal a	7.21e+01
9	57	65.5	935	3	AE-III (peptidylhydro	7.21e+01
10	54	62.1	253	28	Ant(4')-Ia protein.	1.40e+02
11	54	62.1	253	28	Ant(4')-Ia protein.	1.40e+02
12	54	62.1	499	16	Rat cytochrome P450 2	1.40e+02
13	54	62.1	635	38	Green fluorescent pro	1.40e+02
14	53	60.9	494	12	Human steroid-21-hydr	1.75e+02
15	53	60.9	524	27	Arabiopsis thaliana	1.75e+02
16	52	59.8	411	28	Human liver derived m	2.17e+02

17	52	59.8	416	28	W34078	Human liver derived m	2.17e+02
18	52	59.8	419	32	W60149	M. vaccae antigen GV-	2.17e+02
19	52	59.8	461	2	R07450	Rat tumour Necrosis F	2.17e+02
20	52	59.8	508	28	W34075	Human liver derived m	2.17e+02
21	52	59.8	508	26	W16622	Human metalloprotease	2.17e+02
22	52	59.8	980	2	R11113	Equine C-terminal ami	2.17e+02
23	51	58.6	180	8	R50052	ICP34.5 fragment.	2.69e+02
24	51	58.6	241	23	W10016	G-protein coupled rec	2.69e+02
25	51	58.6	328	13	R72984	Epsilon opioid recept	2.69e+02
26	51	58.6	338	4	R21707	HSV-1 (MGH-10) ICP34.	2.69e+02
27	51	58.6	442	31	W46790	Arabiopsis thaliana	2.69e+02
28	51	58.6	564	36	W74581	5' fragment of membra	2.69e+02
29	51	58.6	594	36	W74580	Human membrane protei	2.69e+02
30	51	58.6	594	36	W74579	Rabbit membrane prote	2.69e+02
31	50	57.5	119	15	R80575	Mature Bone Morphogen	3.34e+02
32	50	57.5	594	38	W88552	Secreted protein enco	3.34e+02
33	50	57.5	664	16	R75190	Osteoinductive retrov	3.34e+02
34	50	57.5	1475	2	R08221	Recombinant alpha amy	3.34e+02
35	49	56.3	140	30	W42074	The amino acid sequen	4.13e+02
36	49	56.3	157	21	W20104	H. pylori secreted or	4.13e+02
37	49	56.3	187	35	W71523	Helicobacter polypept	4.13e+02
38	49	56.3	197	22	W20790	H. pylori secreted or	4.13e+02
39	49	56.3	551	27	W26732	Human cytomegalovirus	4.13e+02
40	49	56.3	551	27	W26732	Pseudomonas Orfv secr	4.13e+02
41	49	56.3	877	31	W53828	Pseudomonas Orfv secr	4.13e+02
42	49	56.3	1051	29	W52304	Glucodextranase prote	4.13e+02
43	49	56.3	3391	3	R13166	Proteins encoded by e	4.13e+02
44	49	56.3	3391	23	W06590	Polyprotein of DEN-2	4.13e+02
45	49	56.3	3391	23	W06591	Polyprotein of attenu	4.13e+02

ALIGNMENTS

RESULT 1
ID W37149 standard; Protein; 393 AA.
AC W37149;
DT 06-JUL-1998 (first entry)
DE Murine Ena-Vasp like (Evl) protein.
KW Ena-VASP like protein; Evi protein; Mena; mammalian Ena; Enabled;
KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
KW cell growth; cell motility; mouse.
FH Mus musculus.
FT Key Location/Qualifiers
FT Domain 1..113
FT Binding_site /note= "Ena-VASP homology domain 1 (EVH1)"
FT /note= "mediates binding to profilin"
PN W09801755-A1.
PD 15-JAN-1998.
PF 03-JUL-1997; U11669.
PR 05-JUL-1996; US-675815.
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI Gertler FB, Niebuhr K, Soriano P, Wehland J;
DR WPI: 98-101197/09.
DR N-PSDB: V02997.
PT Detection of modulators of Mena and Ena-VASP-like genes and proteins
PT - used in control of cytoskeletal dynamic events in normal and
PT abnormal cell morphology, adhesion, motility, growth and
PT differentiation
PS Example 1; Page 67-68; 77pp; English.
CC This polypeptide comprises novel murine Ena-VASP like protein Evl.
CC Its amino acid sequence was deduced from a cDNA clone (see V02997)
CC obtained from a mouse embryonic stem cell cDNA library. 2 Novel
CC mammalian genes, mammalian Ena (Mena, see V02996) and Evl, encoding
CC novel proteins Mena (see W37148) and Evl are disclosed. Mena and
CC Evl proteins have a discrete, EVH1 functional domain responsible
CC for Mena binding to Listeria, and to the cytoskeletal proteins
CC zyxin and vinculin. Based on the disclosed Mena and Evl genes and
CC proteins, a variety of methods and compositions are provided for
CC screening, isolating and characterising endogenous and exogenous
CC factors, drugs and therapeutic agents useful to evaluate and/or
CC control cytoskeletal dynamic events involved in normal and abnormal

CC cell morphology, adhesion, motility, growth and/or differentiation.
CC A method of detecting a modulator of Mena activity/expression is
CC claimed.
SQ Sequence 393 AA;

Query Match 75.9%; Score 66; DB 30; Length 393;
Best Local Similarity 80.0%; Pred. No. 9.22e+00;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 80 qwrdrqvyg 89
||||:||||
QY 283 QWRDTR-VYG 291

RESULT 2

ID W37148 standard; Protein: 541 AA.

AC W37148;

DT 06-JUL-1998 (first entry)

DE Mammalian Ena (Mena).

KW Mena protein; mammalian Ena; Enabled; Evi protein; cytoskeleton;

KW cell morphology; cell adhesion; cell differentiation; cell growth;

KW cell motility, mouse.

OS Mus musculus.

FH Key Location/Qualifiers

FT Domain 1..113

FT Peptide /note: "Ena-VASP homology domain 1 (EVH1)"

FT Peptide 156..160

FT Peptide /note: "LERER repeat"

FT Peptide 166..170

FT Peptide /note: "LERER repeat"

FT Peptide 171..175

FT Peptide /note: "LERER repeat"

FT Peptide 176..180

FT Peptide /note: "LERER repeat"

FT Peptide 199..203

FT Peptide /note: "LERER repeat"

FT Peptide 207..211

FT Peptide /note: "LERER repeat"

FT Modified_site 236

FT /note: "O-phosphorylated"

FT Binding_site 305..310

FT /note: "mediates binding to profilin"

FT Binding_site 316..321

FT /note: "mediates binding to profilin"

PN W09801755-A1.

PD 15-JAN-1998.

PF 03-JUL-1997; U11669.

PR 05-JUL-1996; US-675815.

PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.

PI Gertler FB, Niebuhr K, Soriano P, Wehland J;

DR WPI.98-101197/09.

DR N-PSDB; V02996.

PT Detection of modulators of Mena and Ena-VASP-like genes and proteins

PT - used in control of cytoskeletal dynamic events in normal and

PT abnormal cell morphology, adhesion, motility, growth and

PT differentiation

PS Example 1; Page 54-56; 77pp; English.

CC This polypeptide comprises murine Mena (mammalian Ena) that shows

CC significant amino acid similarity to Drosophila Ena and which

CC exhibits a broad pattern of tissue distribution in neurons,

CC fibroblasts, kidney epithelium, muscle, neural crest and

CC haematopoietic cells. Its amino acid sequence was deduced from

CC a cDNA clone (see V02996) obtained from a mouse embryonic stem cell

CC cDNA library. 2 Novel mammalian genes, Mena and Ena-VASP-like

CC (Evi, see V02997), encoding novel proteins Mena and Evi (see

CC W37149) are disclosed. Mena and Evi proteins have a discrete, EVH1

CC functional domain responsible for Mena binding to Listeria, and to

CC the cytoskeletal proteins zyxin and vinculin. The EVH1 domain of

CC Mena is also responsible and sufficient for targeting localisation

CC of Mena and Mena-based fusion proteins to focal adhesions, and to

CC the surface of Listeria cells at the polar site of induction of

CC actin comet tail formation. Based on the disclosed Mena and Evi

CC genes and proteins, a variety of methods and compositions are
CC provided for screening, isolating and characterising endogenous and
CC exogenous factors, drugs and therapeutic agents useful to evaluate
CC and/or control cytoskeletal dynamic events involved in normal and
CC abnormal cell morphology, adhesion, motility, growth and/or
CC differentiation. A method of detecting a modulator of Mena
CC activity/expression is claimed.
SQ Sequence 541 AA;

Query Match 75.9%; Score 66; DB 30; Length 541;
Best Local Similarity 80.0%; Pred. No. 9.22e+00;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 79 qwrdrqvyg 88

||||:||||

QY 283 QWRDTR-VYG 291

RESULT 3

ID W37150 standard; Protein: 740 AA.

AC W37150;

DT 06-JUL-1998 (first entry)

DE Murine Mena, Ena-Vasp like (Evl) protein consensus.

KW Ena-VASP like protein; Evi protein; Mena; mammalian Ena; Enabled;

KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;

KW cell growth; cell motility; mouse.

OS Mus musculus.

FH Key Location/Qualifiers

FT Domain 1..113

FT Misc_difference 8 /note: "Ena-VASP homology domain I (EVH1)"

FT Misc_difference 12 /note: "variable residue"

FT Misc_difference 19..20 /note: "variable residue"

FT Misc_difference 26..29 /note: "variable residues"

FT Misc_difference 30..31 /note: "variable residues"

FT Misc_difference 37 /note: "variable residues"

FT Misc_difference 41..44 /note: "variable residue"

FT Misc_difference 56 /note: "variable residues"

FT Misc_difference 58 /note: "variable residue"

FT Misc_difference 65 /note: "variable residue"

FT Misc_difference 67 /note: "variable residue"

FT Misc_difference 94 /note: "variable residue"

FT Misc_difference 100..101 /note: "variable residues"

FT Misc_difference 104 /note: "variable residue"

FT Misc_difference 107..108 /note: "variable residues"

FT Misc_difference 112 /note: "variable residues"

FT Misc_difference 114..122 /note: "variable residue"

FT Misc_difference 124 /note: "variable residues"

FT Misc_difference 126..138 /note: "variable residue"

FT Misc_difference 140 /note: "variable residues"

FT Misc_difference 145 /note: "variable residue"

FT Misc_difference 148..150 /note: "variable residue"

FT Misc_difference 152 /note= "variable residues"
FT 152 /note= "variable residue"
FT Misc_difference 153..239
FT 153..239 /note= "variable residues"
FT Misc_difference 243
FT 243 /note= "variable residue"
FT Misc_difference 245..413
FT 245..413 /note= "variable residues"
FT Misc_difference 418..420
FT 418..420 /note= "variable residues"
FT Misc_difference 422..423
FT 422..423 /note= "variable residues"
FT Misc_difference 425..433
FT 425..433 /note= "variable residues"
FT Misc_difference 438..444
FT 438..444 /note= "variable residues"
FT Misc_difference 451..454
FT 451..454 /note= "variable residues"
FT Misc_difference 460
FT 460 /note= "variable residue"
FT Misc_difference 462..463
FT 462..463 /note= "variable residues"
FT Misc_difference 466..468
FT 466..468 /note= "variable residues"
FT Misc_difference 470..475
FT 470..475 /note= "variable residues"
FT Misc_difference 477..478
FT 477..478 /note= "variable residues"
FT Misc_difference 488..496
FT 488..496 /note= "variable residues"
FT Misc_difference 498..513
FT 498..513 /note= "variable residues"
FT Misc_difference 519
FT 519 /note= "variable residue"
FT Misc_difference 528..530
FT 528..530 /note= "variable residues"
FT Misc_difference 532
FT 532 /note= "variable residue"
FT Misc_difference 535..562
FT 535..562 /note= "variable residues"
FT Misc_difference 565
FT 565 /note= "variable residue"
FT Misc_difference 572..574
FT 572..574 /note= "variable residues"
FT Misc_difference 580..592
FT 580..592 /note= "variable residues"
FT Misc_difference 594..601
FT 594..601 /note= "variable residues"
FT Misc_difference 603..620
FT 603..620 /note= "variable residues"
FT Misc_difference 622..623
FT 622..623 /note= "variable residues"
FT Misc_difference 625
FT 625 /note= "variable residue"
FT Misc_difference 629..630
FT 629..630 /note= "variable residues"
FT Misc_difference 633..637
FT 633..637 /note= "variable residues"
FT Misc_difference 639..641
FT 639..641 /note= "variable residues"
FT Misc_difference 643..702
FT 643..702 /note= "variable residues"
FT Misc_difference 707
FT 707 /note= "variable residue"
FT Misc_difference 713
FT 713 /note= "variable residue"
FT Misc_difference 715..717
FT 715..717 /note= "variable residues"
FT Misc_difference 720
FT 720 /note= "variable residue"
FT Misc_difference 731
FT 731 /note= "variable residue"

FT Misc_difference 735..740
FT 735..740 /note= "variable residues"
FT Domain
FT 507..733 /note= "C-terminal homology domain"
FT WO9801755-AL.
FT 15-JAN-1998.
FT 03-JUL-1997; U11669.
FT 05-JUL-1996; US-675815.
FT (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
FT (HUTC-) HUTCHINSON CANCER RES CENT FRED.
FT Gertler FB, Niebuhr K, Soriano P, Wehland J;
FT WPI; 98-101197/09.
FT Detection of modulators of Mena and Ena-VASP-like genes and proteins
FT - used in control of cytoskeletal dynamic events in normal and
FT abnormal cell morphology, adhesion, motility, growth and
FT differentiation
FT Example 1; Page 68-70; 77pp; English.
FT This polypeptide comprises a consensus of novel murine mammalian Ena
FT (Mena) protein (see W37148) and novel murine Ena-VASP like protein
FT Evi (see W37149). The greatest degree of amino acid identity is
FT in the N-terminal Ena-VASP homology domain 1 that is responsible
FT for Mena binding to Listeria, and to the cytoskeletal proteins
FT CC xilin and vinculin. The C-terminal homology domain contains a
FT CC putative G-actin binding sequence and a conserved charge cluster.
FT CC Based on the disclosed Mena and Evi genes and proteins, a variety
FT CC of methods and compositions are provided for screening, isolating
FT CC and characterising endogenous and exogenous factors, drugs and
FT CC therapeutic agents useful to evaluate and/or control cytoskeletal
FT CC dynamic events involved in normal and abnormal cell morphology,
FT CC adhesion, motility, growth and/or differentiation. A method of
FT CC detecting a modulator of Mena activity/expression is claimed.
FT Sequence 740 AA:
SQ
Query Match 75.9%; Score 66; DB 30; Length 740;
Best Local Similarity 80.0%; Pred. No. 9.22e+00;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
Db 81 qvrdarqvvg 90
||||| |||
QY 283 QWRDTR-VYG 291
RESULT 4
ID W37151 standard; Protein; 783 AA.
AC W37151;
DT 06-JUL-1998 (first entry)
DE Mouse neural Mena+ protein.
KW Neural Mena+ protein; mammalian Ena; Enabled protein; Evi protein;
KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
KW cell growth; cell motility; mouse.
OS Mus musculus.
FH Key Location/Qualifiers
FT Misc_difference 378 /note= "encoded by GGN"
FT WO9801755-AL.
FT 15-JAN-1998.
FT 03-JUL-1997; U11669.
FT 05-JUL-1996; US-675815.
FT (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
FT (HUTC-) HUTCHINSON CANCER RES CENT FRED.
FT Gertler FB, Niebuhr K, Soriano P, Wehland J;
FT WPI; 98-101197/09.
FT N-PSDB; V02998.
FT Detection of modulators of Mena and Ena-VASP-like genes and proteins
FT - used in control of cytoskeletal dynamic events in normal and
FT abnormal cell morphology, adhesion, motility, growth and
FT differentiation
FT Example 4; Page 58-60; 77pp; English.
FT This protein comprises novel murine neural Mena+. Its amino acid
FT sequence was deduced from a cDNA clone (see V02998) obtained from
FT a mouse brain cDNA library. Neural Mena+ contains an exon that
FT introduces 244 amino acids between amino acids 238 and 239 of
FT mammalian Ena (Mena, see W37148). Two other isoforms, neural

CC Mena++ (see W37152) and neural Mena+++ (see W37153), are also
 CC disclosed. Unlike Mena, neural Mena isoforms exhibit neural
 CC tissue-specific distribution. Based on the disclosed Mena and Evi
 CC genes (see also V02996-97) and proteins (see also W37148-49), a
 CC variety of methods and compositions are provided for screening,
 CC isolating and characterising endogenous and exogenous factors,
 CC drugs and therapeutic agents useful to evaluate and/or control
 CC cytoskeletal dynamic events involved in normal and abnormal cell
 CC morphology, adhesion, motility, growth and/or differentiation. A
 CC method of detecting a modulator of Mena activity/expression is
 CC claimed.
 SQ Sequence 783 AA;

Query Match 75.9%; Score 66; DB 30; Length 783;
 Best Local Similarity 80.0%; Pred. No. 9.22e+00;
 Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 79 qwrdrqvyg 88
 ||||| |||
 QY 283 QWRDTR-VYG 291

RESULT 5

ID W37152 standard; Protein; 787 AA.
 AC W37152;
 DT 06-JUL-1998 (first entry)
 DE Mouse neural Mena++ protein.
 KW Neural Mena++ protein; mammalian Ena; Enabled protein; Evi protein;
 KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
 KW cell growth; cell motility; mouse.
 OS Mus musculus.
 PN W09801755-A1.
 PD 15-JAN-1998.
 PF 03-JUL-1997; U11669.
 PR 05-JUL-1996; US-675815.
 PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
 PI Gertler FB, Niebuhr K, Soriano P, Wehland J;
 DR WPI: 98-101197/09.
 PT Detection of modulators of Mena and Ena-VASP-like genes and proteins
 PT - used in control of cytoskeletal dynamic events in normal and
 PT abnormal cell morphology, adhesion, motility, growth and
 PT differentiation

PS Example 4; Page 60-63; 77pp; English.
 CC This protein comprises novel murine neural Mena++. Its amino acid
 CC sequence was deduced from a cDNA clone obtained from a mouse brain
 CC cDNA library. Two other isoforms, neural Mena+ (see W37151) and
 CC neural Mena+++ (see W37153), are also disclosed. Unlike mammalian
 CC Ena (Mena, see W37148), neural Mena isoforms exhibit neural
 CC tissue-specific distribution. Based on the disclosed Mena and Evi
 CC genes (see also V02996-98) and proteins (see also W37148-49), a
 CC variety of methods and compositions are provided for screening,
 CC isolating and characterising endogenous and exogenous factors,
 CC drugs and therapeutic agents useful to evaluate and/or control
 CC cytoskeletal dynamic events involved in normal and abnormal cell
 CC morphology, adhesion, motility, growth and/or differentiation. A
 CC method of detecting a modulator of Mena activity/expression is
 CC claimed.
 SQ Sequence 787 AA;

Query Match 75.9%; Score 66; DB 30; Length 787;
 Best Local Similarity 80.0%; Pred. No. 9.22e+00;
 Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 79 qwrdrqvyg 88
 ||||| |||
 QY 283 QWRDTR-VYG 291

RESULT 6

ID W37153 standard; Protein; 802 AA.
 AC W37153;
 DT 06-JUL-1998 (first entry)

DE Mouse neural Mena+++ protein.
 KW Neural Mena+++ protein; mammalian Ena; Enabled protein; Evi protein;
 KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
 KW cell growth; cell motility; mouse.
 OS Mus musculus.
 PN W09801755-A1.
 PD 15-JAN-1998.
 PF 03-JUL-1997; U11669.
 PR 05-JUL-1996; US-675815.
 PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
 PI Gertler FB, Niebuhr K, Soriano P, Wehland J;
 DR WPI: 98-101197/09.
 PT Detection of modulators of Mena and Ena-VASP-like genes and proteins
 PT - used in control of cytoskeletal dynamic events in normal and
 PT abnormal cell morphology, adhesion, motility, growth and
 PT differentiation

PS Example 4; Page 63-65; 77pp; English.

CC This protein comprises novel murine neural Mena++. Its amino acid
 CC sequence was deduced from a cDNA clone obtained from a mouse brain
 CC cDNA library. Two other isoforms, neural Mena+ (see W37151) and
 CC neural Mena++ (see W37152), are also disclosed. Unlike mammalian
 CC Ena (Mena, see W37148), neural Mena isoforms exhibit neural
 CC tissue-specific distribution. Based on the disclosed Mena and Evi
 CC genes (see also V02996-98) and proteins (see also W37148-49), a
 CC variety of methods and compositions are provided for screening,
 CC isolating and characterising endogenous and exogenous factors,
 CC drugs and therapeutic agents useful to evaluate and/or control
 CC cytoskeletal dynamic events involved in normal and abnormal cell
 CC morphology, adhesion, motility, growth and/or differentiation. A
 CC method of detecting a modulator of Mena activity/expression is
 CC claimed.
 SQ Sequence 802 AA;

Query Match 75.9%; Score 66; DB 30; Length 802;
 Best Local Similarity 80.0%; Pred. No. 9.22e+00;
 Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 79 qwrdrqvyg 88
 ||||| |||
 QY 283 QWRDTR-VYG 291

RESULT 7

ID P94854 standard; protein; 875 AA.
 AC P94854;
 DT 27-JUN-1990 (first entry)
 DE C-terminal prepro-C-terminal alpha-amidating enzyme of pXA799.
 KW alpha-amidating; pAX799; alpha amide; ds.
 OS Synthetic.

PN EP-299790-A.
 PD 18-JAN-1989.
 PF 15-JUL-1988; 306508.
 PR 17-JUL-1987; JP-177184.
 PR 05-DEC-1987; JP-306867.
 PA (SUNR) Suntory Ltd.
 PI Ohnuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
 DR WPI: 89-017279/03.
 DR N-P5DB; N90791.

PT Recombinant C terminal alpha amidating enzymes of Xenopus laevis -
 PT and their precursors deoxyribonucleic acid encoding sequences.
 PS Disclosure; 7pp; English.
 CC Plasmid pXA799 contains a sequence derived from Xenopus laevis.
 CC The plasmid was screened from an E.coli library using plasmid pXA457
 CC to screen a larger library.
 CC Although the gene product is similar to that of pXA457 at the N-terminus,
 CC it has an area of hydrophobic elements suggesting a membrane function.
 CC See also N93060.
 SQ Sequence 875 AA;

Query Match 65.5%; Score 57; DB 1; Length 875;
 Best Local Similarity 55.6%; Pred. No. 7.21e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

Db 788 rwrkvrmyg 796
QY 283 QWRDTRVYG 291

      111 1:11
      283 QWRDTRVYG 291

RESULT 8
ID R73053 standard; Protein; 875 AA.
AC R73053.
DE 06-NOV-1995 (first entry)
KW Peptidyl C-terminal alpha-amidating enzyme.
KW Trichostatin; CHO.
OS Not specified.
FH Key Location/Qualifiers
FT peptide 1..39
FT /label= Sig_peptide
FN EP-649900-A.
PD 26-APR-1995.
PF 07-SEP-1994; 306587.
PR 08-SEP-1993; JP-257881.
PA (FURU/) FURUKAWA K.
PI Furukawa K, Ohsuye K, Sugimura K;
DR WPI: 95-156754/21.
DR N-PSDB: Q87970.
PT Increasing protein prodn. from cultured animal cells - by adding
PT a trichostatin to the medium, effective at low concn. and not
PT injurious to host cells
PS Disclosure: Page 10-15; 19pp; English.
CC 3mu-15 cells (i.e. CHO cells expressing a peptidyl C-terminal alpha-
CC amidase enzyme (AE)) were suspended in F-12 medium to which aliquots
CC of trichostatin were added. Cells were cultured for 3 days at 37 deg
CC and then assayed for AE. Without trichostatin, AE productivity was
CC 708 U/ml. At 10, 50, 100 and 200 nM trichostatin, AE productivities
CC were respectively 866, 1897, 1894 and 3359 U/ml.
SQ Sequence 875 AA;

      Query Match 65.5%; Score 57; DB 13; Length 875;
      Best Local Similarity 55.6%; Pred. No. 7.21e+01;
      Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 788 rwrkvrmyg 796
QY 283 QWRDTRVYG 291

      111 1:11
      283 QWRDTRVYG 291

RESULT 9
ID R20112 standard; Protein; 935 AA.
AC R20112;
DE 06-APR-1992 (first entry)
DE AE-III (peptidylhydroxyglycine N-C lyase precursor).
KW Amidation; PHL.
OS Xenopus laevis.
FH Key Location/Qualifiers
FT protein 1..935
FT /label= AE-III
FT /note= "including PAM and PHL domains"
FT region 383..935
FT /*tag= b
FT /product= PHL
FN EP-465404-A.
PD 08-JAN-1992.
PF 27-MAY-1991; 810399.
PR 01-JUN-1990; JP-141678.
PR 10-AUG-1990; JP-210535.
PR 30-NOV-1990; JP-329911.
PA (CIBA ) CIBA GEIGY AG.
PI Iwasaki Y, Shimoi H, Suzuki K, Ghisalba D, Nishikawa Y;
PI Kawahara T, Kangawa K;
DR WPI: 92-010570/02.
DR N-PSDB: Q20269.
PT Novel DNA encoding peptidyl hydroxyglycine N-C lyase (PHL) -

      Query Match 62.1%; Score 54; DB 28; Length 253;
      Best Local Similarity 60.0%; Pred. No. 1.40e+02;
      Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 148 gkwrnrvqg 157
QY 282 QWRDTRVYG 291

      111 1:11
      282 QWRDTRVYG 291

RESULT 11
ID W31402 standard; Protein; 253 AA.
AC W31402;
DT 11-MAY-1998 (first entry)

      Query Match 65.5%; Score 57; DB 3; Length 935;
      Best Local Similarity 55.6%; Pred. No. 7.21e+01;
      Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 847 rwrkvrmyg 855
QY 283 QWRDTRVYG 291

      111 1:11
      283 QWRDTRVYG 291

RESULT 10
ID W31408 standard; Protein; 253 AA.
AC W31408;
DT 11-MAY-1998 (first entry)
DE Ant(4')-Ia protein.
KW Kanamycin resistance gene; adenyl 4'-nucleotidyltransferase type Ia;
KW ant(4')-Ia gene; composite nucleic acid molecule synthesis;
KW site specific DNA fragment ligation; PCR product ligation.
OS Synthetic.
OS Escherichia coli.
PN WO9742330-A1.
PD 13-NOV-1997.
PR 06-MAY-1997; U07698.
PR 06-MAY-1996; US-642045.
PA (APOL-) APOLLON INC.
PI Pachuk CJ, Samuel M, Satishchandran C, Zurawski JA;
DR WPI: 97-558988/51.
DR N-PSDB: V02474.
PT Synthesis of composite nucleic acid molecules - by chain reaction
PT cloning using nucleic acid molecules, a bridging oligo-nucleotide
PT and a thermostable ligase
PS Disclosure: Page 40; 69pp; English.
CC This sequence represents the kanamycin resistance protein adenyl
CC 4'-nucleotidyltransferase type Ia (ant(4')-Ia). DNA encoding this
CC sequence is stated in the specification as being chimeric. The promoter
CC and terminator regions are stated as being from a different gene, however
CC no indication is given as to which portions of this sequence are from the
CC other gene, or which organism the other gene is from. This sequence can
CC be synthesised using the method of the invention, which is for
CC synthesising a composite nucleic acid molecule (NAM) by ligating a first
CC NAM to a second NAM. The methods provide for the generation of larger
CC NAMs from smaller ones without the need to treat the ends in such a way
CC to ensure specific binding of two molecules in correct orientation. They
CC can be used for the site specific ligation of DNA fragments generated by
CC restriction enzyme digestion, DNase digestion, chemical cleavage,
CC enzymatic or chemical synthesis, particularly for ligation of PCR
CC products.
SQ Sequence 253 AA;

      Query Match 62.1%; Score 54; DB 28; Length 253;
      Best Local Similarity 60.0%; Pred. No. 1.40e+02;
      Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 148 gkwrnrvqg 157
QY 282 QWRDTRVYG 291

      111 1:11
      282 QWRDTRVYG 291

RESULT 11
ID W31402 standard; Protein; 253 AA.
AC W31402;
DT 11-MAY-1998 (first entry)

```

PT used to prepare PHL which can be used in the amidation of
PT peptide(s) e.g. human calcitonin.
PS Claim 4; Page 18; 28pp; English.

CC The sequence was deduced from a cDNA insert from pAE-III-202-4
CC (FERM BP-3172). The vector serves as a source for a DNA fragment
CC encoding PHL for the construction of an expression vector for the
CC prepn. of recombinant PHL. The protein may be truncated to
CC comprise only residues 383-706 or 383-713. The PHL catalyses the
CC reaction: R-GlyOH -> R-NH2. It can be used to produce peptides with
CC amidated C-termini, e.g. calcitonin, growth hormone, LH-RH.
SQ Sequence 935 AA;

Query Match 65.5%; Score 57; DB 3; Length 935;
Best Local Similarity 55.6%; Pred. No. 7.21e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 847 rwrkvrmyg 855
QY 283 QWRDTRVYG 291
111 1:11

RESULT 10
ID W31408 standard; Protein; 253 AA.
AC W31408;
DT 11-MAY-1998 (first entry)
DE Ant(4')-Ia protein.
KW Kanamycin resistance gene; adenyl 4'-nucleotidyltransferase type Ia;
KW ant(4')-Ia gene; composite nucleic acid molecule synthesis;
KW site specific DNA fragment ligation; PCR product ligation.
OS Synthetic.
OS Escherichia coli.
PN WO9742330-A1.
PD 13-NOV-1997.
PR 06-MAY-1997; U07698.
PR 06-MAY-1996; US-642045.
PA (APOL-) APOLLON INC.
PI Pachuk CJ, Samuel M, Satishchandran C, Zurawski JA;
DR WPI: 97-558988/51.
DR N-PSDB: V02474.

PT Synthesis of composite nucleic acid molecules - by chain reaction
PT cloning using nucleic acid molecules, a bridging oligo-nucleotide
PT and a thermostable ligase
PS Disclosure: Page 40; 69pp; English.

CC This sequence represents the kanamycin resistance protein adenyl
CC 4'-nucleotidyltransferase type Ia (ant(4')-Ia). DNA encoding this
CC sequence is stated in the specification as being chimeric. The promoter
CC and terminator regions are stated as being from a different gene, however
CC no indication is given as to which portions of this sequence are from the
CC other gene, or which organism the other gene is from. This sequence can
CC be synthesised using the method of the invention, which is for
CC synthesising a composite nucleic acid molecule (NAM) by ligating a first
CC NAM to a second NAM. The methods provide for the generation of larger
CC NAMs from smaller ones without the need to treat the ends in such a way
CC to ensure specific binding of two molecules in correct orientation. They
CC can be used for the site specific ligation of DNA fragments generated by
CC restriction enzyme digestion, DNase digestion, chemical cleavage,
CC enzymatic or chemical synthesis, particularly for ligation of PCR
CC products.
SQ Sequence 253 AA;

Query Match 62.1%; Score 54; DB 28; Length 253;
Best Local Similarity 60.0%; Pred. No. 1.40e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 148 gkwrnrvqg 157
QY 282 QWRDTRVYG 291
111 1:11

RESULT 11
ID W31402 standard; Protein; 253 AA.
AC W31402;
DT 11-MAY-1998 (first entry)

DE Ant(4')-Ia protein.
 KW Adenylyl 4'-nucleotidyl transferase type Ia: ant(4')-Ia gene: aph(3')-Ia;
 KW chimeric gene; aminoglycoside 3'-phosphotransferase type Ia;
 KW promoter region; kanamycin resistance gene; antigen delivery vector;
 KW gene therapy; genetic immunisation; aminoglycoside.
 OS Chimeric: Escherichia coli.
 OS Chimeric: Staphylococcus aureus.
 PN WO9741892-A1.
 PD 13-NOV-1997.
 PR 06-MAY-1997; U07853.
 PR 06-MAY-1996; US-642045.
 PA (APOL-) APOLLON INC.
 PI Satishchandran C, Snyder LA;
 DR WPI: 97-558697/51.
 DR N-PSDB: V02453.
 PT Chimeric kanamycin resistance gene - comprising sequence encoding
 PT adenylyl 4'-nucleotidyl-transferase type Ia and heterologous
 PT promoter and terminator
 PS Example 1: Page 32-33; 51pp; English.
 CC This sequence is encoded by a chimeric gene containing the Staphylococcus
 CC aureus adenylyl 4'-nucleotidyl transferase type Ia (ant(4')-Ia) open
 CC reading frame attached to the E coli aminoglycoside 3'-phosphotransferase
 CC type Ia (aph(3')-Ia) promoter region. The DNA is an example of the
 CC chimeric kanamycin resistance (CKR) gene of the invention, which
 CC comprises a ant(4')-Ia coding sequence, operably linked to a heterologous
 CC promoter and termination sequence. The CKR gene can be used in the
 CC production of protein or antigen delivery vectors, e.g. for therapeutic
 CC or prophylactic treatment, gene therapy or genetic immunisation.
 CC Ant(4')-Ia confers resistance to a much more limited number of clinically
 CC relevant aminoglycosides compared to aph(3')-Ia, and therefore provides
 CC clinical vectors with improved safety.
 SQ Sequence 253 AA;

Query Match 62.1%; Score 54; DB 28; Length 253;
 Best Local Similarity 60.0%; Pred. No. 1.40e+02;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 148 gkwnrivrvg 157
 ||: ||:
 QY 282 GQWRDTRVYG 291

RESULT 12
 ID R94349 standard; Protein: 499 AA.
 AC R94349;
 DT 20-JUN-1996 (first entry)
 DE Rat cytochrome P450 2C11.
 KW Cytochrome P450 2C11; arachidonic acid epoxidase;
 KW salt-induced hypertension; diagnosis; transgenic animal;
 KW therapy.
 OS Rattus sp.
 PN WO9610074-A1.
 PD 04-APR-1996.
 PF 27-SEP-1995; U13051.
 PR 28-SEP-1994; US-314601.
 PA (UYVA-) UNIV VANDERBILT.
 PI Capdevilla J, Karara A, Makita K;
 DR WPI: 96-200913/20.
 PT Diagnosis and treatment of salt-induced hypertension - using prods.
 PT developed using mutant P450 2C11 arachidonic acid epoxidase gene.
 PS Disclosure: Page 40-42; 53pp; English.
 CC Normally expressed rat cytochrome P450
 CC epoxidase (PAAE) has the amino acid sequence given in R94349.
 CC Mutations in PAAE and its human homologue are associated with a
 CC predisposition to salt-induced hypertension (SIH). Products and
 CC methods have been developed that can be used in the diagnosis and
 CC treatment of SIH, in screening assays for effective drugs, and in the
 CC breeding of transgenic animals.
 CC N.B. the translated sequence of the PAAE gene (T13412) given in
 CC the specification does not correspond to this rat PAAE amino acid
 CC sequence. 499 AA;
 SQ

Query Match 62.1%; Score 54; DB 16; Length 499;
 Best Local Similarity 40.0%; Pred. No. 1.40e+02;
 Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 117 gqwkairrfs 126
 |||:: ||:
 QY 282 GQWRDTRVYG 291

RESULT 13
 ID W85034 standard; Protein: 635 AA.
 AC W85034;
 DT 08-FEB-1999 (first entry)
 DE Green fluorescent protein-VASP fusion product.
 KW Human; VASP gene; fusion protein; green fluorescent protein; GFP;
 KW intracellular signalling; chimera.
 OS Chimeric - Aequorea victoria.
 OS Chimeric - Homo sapiens.
 PN WO9845704-A2.
 PD 15-OCT-1998.
 PF 07-APR-1998; DK0145.
 PR 07-APR-1997; DK-000392.
 PA (NOVO) NOVO-NORDISK AS.
 PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,
 PI Tullin S;
 DR WPI: 98-594491/50.
 DR N-PSDB: V71079.
 PT Determining effect on signalling pathways in live cells from
 PT redistribution of luminophores - specifically fusions of green
 PT fluorescent protein with a signalling component, and new apparatus,
 PT particularly for identifying toxins and potential therapeutic agents
 PS Example 6; Pages 244-245; 326pp; English.
 CC The present sequence represents a green fluorescent protein (GFP)-human
 CC VASP fusion protein. The fusion protein is used in an assay to
 CC exemplify the invention. The specification describes how quantitative
 CC information about the influence of a molecule on a cellular response is
 CC obtained by recording the variation, caused by the molecule, on
 CC mechanically intact living cells, in the spatially distributed light
 CC emitted from a luminophore present in the cells. The variation in light
 CC emission is processed to provide information that correlates spatial
 CC distribution to the degree of the molecule. The method is used to
 CC identify agents that (in)directly affect intracellular signalling,
 CC especially to screen for potential therapeutic agents or toxins, and to
 CC identify new drug targets.
 SQ Sequence 635 AA;

Query Match 62.1%; Score 54; DB 38; Length 635;
 Best Local Similarity 70.0%; Pred. No. 1.40e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Db 336 qwrdratqvg 345
 ||||: ||:
 QY 283 QWRDTR-VYG 291

RESULT 14
 ID R62825 standard; Protein: 494 AA.
 AC R62825;
 DT 07-JUN-1995 (first entry)
 DE Human steroid-21-hydroxylase.
 KW Steroid-21-hydroxylase; 21-OH; adrenal autoantibody; epitope;
 KW autoimmune disease; Addison disease.
 OS Homo sapiens.
 FH key Location/Qualifiers
 FT region 227..480
 FT /note= "epitopic region for adrenal autoantibody"
 FT WO9425604-A.
 PN 10-NOV-1994.
 PD 22-APR-1994; G00855.
 PR 24-APR-1993; GB-008544.
 PA (RSRR-) RSR LTD.
 PI Furmaniak-Wehr JM;
 DR WPI: 94-358280/44.

DR N-PSDB; Q73170.
PT New steroid 21-hydroxylase fragments - contg. an epitope region
PT for adrenal auto:antibody used in diagnosis and therapy of
PT adrenal auto:immune disease
PS Disclosure; Page 20-22; 37pp; English.
CC Human steroid-21-hydroxylase (21-OH) includes an epitope region for
CC adrenal autoantibody (AA) useful for treatment of e.g. Addison
CC disease. The epitopic region was identified from a series of
CC truncations/deletions 3' from a leader sequence, expression in
CC Saccharomyces cerevisiae, and testing for binding rabbit antibody
CC to 21-OH and AA.
SQ Sequence 494 AA;

Query Match 60.9%; Score 53; DB 12; Length 494;
Best Local Similarity 66.7%; Pred. No. 1.75e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 251 gqwrmdmndy 259
|||||
Qy 282 GOWRDRVY 290

RESULT 15

ID W35346 standard; Protein: 524 AA.
AC W35346;
DT 14-APR-1998 (first entry)
DE Arabidopsis thaliana epsilon cyclase.
KW Epsilon cyclase; E-cyclase; carotenoid; biosynthetic enzyme;
KW pigment; vector; pAREps.
OS Arabidopsis thaliana.
PN W09736998-A1.
PD 09-OCT-1997.
PF 28-JAN-1997; U00540.
PR 29-MAR-1996; US-624125.
PA (UYMA-) UNIV MARYLAND BALTIMORE.
PI Cunninghamham FX, Sun Z;
DR WPI: 97-503091/46.
DR N-PSDB; T95371.
PT Eukaryotic carotenoid biosynthetic enzymes and related genes -
PT useful to control ratio of various carotenoid(s) in host and for
PT production of novel carotenoid pigments
PS Claim 1; Page 33-35; 89pp; English.
CC This protein comprises Arabidopsis thaliana epsilon cyclase, an
CC enzyme responsible for the formation of epsilon end-groups in
CC carotenoids. Its amino acid sequence was deduced from an
CC isolated cDNA clone (see T95371). Expression vector pAREps
CC comprising the epsilon cyclase gene is deposited as ATCC 98005.
CC The claimed eukaryotic carotenoid biosynthetic enzymes
CC epsilon cyclase, beta-carotene hydroxylase and isopenentenyl
CC pyrophosphate isomerase (see W35346-51) are used in methods for
CC augmenting the accumulation of carotenoids and for the production
CC of novel and rare carotenoids in host cells. Methods are also
CC provided for controlling the ratio of various carotenoids in a
CC host, and for screening for eukaryotic genes that encode enzymes
CC of carotenoid biosynthesis and metabolism.
SQ Sequence 524 AA;

Query Match 60.9%; Score 53; DB 27; Length 524;
Best Local Similarity 71.4%; Pred. No. 1.75e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 167 wrctivy 173
|||||
Qy 284 WRDRVY 290

Search completed: Thu Jul 8 18:46:14 1999
Job time : 25 secs.

THIS PAGE BLANK (USPTO)

M P S R L H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:45:15 1999; MasPar time 5.02 Seconds
Tabular output not generated. 79.860 Million cell updates/sec

Title: >US-09-041-236-2
Description: (282-291) from US09041236.pep (19 of 45)
Perfect Score: 87
Sequence: 1 GQWRDTRVYG 10

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 25.854; Variance 40.694; scale 0.635

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description Pred. No.
1	66	75.9	393 2	JC5614 RNB6 protein - rat 2.97e-01
2	61	70.1	684 2	A56154 Abl substrate ena (en 2.20e+00
3	57	65.5	179 2	S22531 pathogenesis-related 1.02e+01
4	57	65.5	196 2	A35177 chromate resistance p 1.02e+01
5	57	65.5	875 1	URXLA2 peptidylglycine monoo 1.02e+01
6	57	65.5	935 2	SI7855 peptidylglycine monoo 1.02e+01
7	56	64.4	312 2	S77365 hypothetical protein 1.49e+01
8	56	64.4	434 2	A54588 protein phosphatase A 1.49e+01
9	56	64.4	686 2	C64428 hypothetical protein 1.49e+01
10	56	64.4	878 2	C71305 probable leucyl-tRNA 1.49e+01
11	55	63.2	202 2	H70544 hypothetical protein 2.15e+01
12	55	63.2	2276 2	T00076 hypothetical protein 2.15e+01
13	55	63.2	2554 1	TVFF7L kinase-related protei 2.15e+01
14	54	62.1	125 1	ERBP4Z gene 4 protein - phag 3.10e+01
15	54	62.1	125 1	ERBP4Z early protein gp4 - p 3.10e+01
16	54	62.1	253 2	B24456 kanamycin nucleotidyl 3.10e+01
17	54	62.1	253 2	A24456 kanamycin nucleotidyl 3.10e+01
18	54	62.1	256 2	S09565 neomycin resistance p 3.10e+01
19	54	62.1	273 2	E71913 utp--glucose-1-phosph 3.10e+01
20	54	62.1	384 2	S51797 vasodilator-stimulate 3.10e+01
21	54	62.1	384 2	S51796 vasodilator-stimulate 3.10e+01
22	54	62.1	437 2	S34469 gamma-aminobutyric ac 3.10e+01
23	53	60.9	53 2	SI7672 hypothetical protein 4.45e+01

24	53	60.9	203	2	S76996	hypothetical protein	4.45e+01
25	53	60.9	494	1	O4HUC2	steroid 21-monooxygen	4.45e+01
26	53	60.9	497	1	A43349	steroid 21-monooxygen	4.45e+01
27	53	60.9	499	1	B43684	RNA-directed RNA poly	4.45e+01
28	53	60.9	536	2	S56542	thioglycosidase (EC 3	4.45e+01
29	53	60.9	1091	2	E71322	probable isoleucyl-tr	4.45e+01
30	53	60.9	1432	2	S58819	antiviral protein SKI	4.45e+01
31	53	60.9	1452	1	SI7870	protein-tyrosine-phos	4.45e+01
32	53	60.9	1452	1	SI7869	protein-tyrosine-phos	4.45e+01
33	53	60.9	2638	1	A42545	genome polypotein -	4.45e+01
34	53	60.9	3412	1	GNWVVB	genome polypotein -	4.45e+01
35	53	60.9	3430	1	GNWVWV	genome polypotein -	4.45e+01
36	53	60.9	3433	1	GNWVKV	genome polypotein -	4.45e+01
37	52	59.8	144	2	I70695	omega protein - human	6.35e+01
38	52	59.8	213	1	A33911	pre-B cell ig lambda-	6.35e+01
39	52	59.8	256	2	S63371	probable membrane pro	6.35e+01
40	52	59.8	419	2	A37926	fumarylacetoacetase (6.35e+01
41	52	59.8	496	1	O4BOC2	steroid 21-monooxygen	6.35e+01
42	52	59.8	508	2	JC5082	matrix metalloprotein	6.35e+01
43	52	59.8	531	2	S76041	hypothetical protein	6.35e+01
44	52	59.8	632	2	H70339	NADH dehydrogenase I	6.35e+01
45	52	59.8	777	2	G69310	penicillin G acylase	6.35e+01

ALIGNMENTS

RESULT 1

ENTRY JC5614 #type complete
TITLE RNB6 protein - rat
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 23-Sep-1997 #sequence_revision 17-Oct-1997 #text_change 17-Mar-1999

ACCESSIONS JC5614
REFERENCE JC5614
#authors Ohta, S.; Mineta, T.; Kimoto, M.; Tabuchi, K.
#journal Biochem. Biophys. Res. Commun. (1997) 237:307-312
#title Differential display cloning of a novel rat cDNA (RNB6) that shows high expression in the neonatal brain revealed a member of Ena/VASP family.

#cross-references MUID:97415794
#accession JC5614
#molecule_type mRNA
##residues 1-393 ##label OHT
##cross-references GB:070211; NID:g2058461; PID:g2058462
##experimental_source brain

COMMENT This protein belongs to Ena/VASP family member, and is involved in the development of central nervous system through the control of neural cell motility and including neuronal fiber extension.

SUMMARY #length 393 #molecular-weight 42095 #checksum 9628

Query Match 75.9% Score 66; DB 2; Length 393;
Best Local Similarity 80.0%; Pred. No. 2.97e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 80 QWRDARQVYG 89
|||||
QY 283 QWRDTR-VYG 291

RESULT 2

ENTRY A56154 #type complete
TITLE Abl substrate ena (enabled) - fruit fly (Drosophila melanogaster)
ORGANISM #formal name Drosophila melanogaster
DATE 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 24-Sep-1998

ACCESSIONS A56154
REFERENCE A56154
#authors Gertler, F.B.; Comer, A.R.; Juang, J.L.; Ahern, S.M.; Clark, M.J.; Liebli, E.C.; Hoffmann, F.M.
#journal Genes Dev. (1995) 9:521-533
#title Drosophila Abl tyrosine kinase, encodes an Abl substrate

```
#cross-references MUID:95212903
#accession A56154
#status preliminary
#molecule_type DNA
#residues 1-684 #label GER
##cross-references GB:021123; NID:g758428; PID:g755821
GENETICS
#gene
#ena
##cross-references FlyBase:FBgn0000578
phosphoprotein
SUMMARY
#length 684 #molecular-weight 72049 #checksum 8892
Query Match 70.1%; Score 61; DB 2; Length 684;
Best Local Similarity 70.0%; Pred. No. 2.20e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 1; Gaps 1;
Db 79 QWRDSRFVYG 88
||||: |||
QY 283 QWRDTR-VYG 291

RESULT 3
ENTRY S22531 #type complete
TITLE pathogenesis-related protein 1b - common tobacco
ORGANISM #formal_name Nicotiana tabacum #common_name common tobacco
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
08-Sep-1997
ACCESSIONS S22531
REFERENCE S22531
#authors Eyal, Y.; Sagee, O.; Fluhr, R.
#journal Plant Mol. Biol. (1992) 19:589-599
#title Dark-induced accumulation of a basic pathogenesis-related
(PR-1) transcript and a light requirement for its induction
by ethylene.
#accession S22531
##molecule_type DNA
##residues 1-179 #label EYA
##cross-references EMBL:X66942; NID:g19969; PID:g19970
GENETICS
#gene prb-1b
CLASSIFICATION #superfamily pathogenesis-related leaf protein
SUMMARY #length 179 #molecular-weight 20120 #checksum 711
Query Match 65.5%; Score 57; DB 2; Length 179;
Best Local Similarity 60.0%; Pred. No. 1.02e+01;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 152 GNRGRRPYG 161
||||: |||
QY 282 GQWRDTRVYG 291

RESULT 4
ENTRY A35177 #type complete
TITLE chromate resistance protein B - Alcaligenes eutrophus plasmid
PMOL28
ORGANISM #formal_name Alcaligenes eutrophus
DATE 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change
09-Sep-1997
ACCESSIONS A35177
REFERENCE A35177
#authors Nies, A.; Nies, D.H.; Silver, S.
#journal J. Biol. Chem. (1990) 265:5648-5653
#title Nucleotide sequence and expression of a plasmid-encoded
chromate resistance determinant from Alcaligenes eutrophus.
#cross-references MUID:9020806
#accession A35177
#status preliminary
#molecule_type DNA
#residues 1-196 #label NIE
##cross-references GB:J05278; NID:g141911; PID:g141912
GENETICS

with SH3 domain-binding properties.
#cross-references MUID:95212903
#accession A56154
#status preliminary
#molecule_type DNA
#residues 1-684 #label GER
##cross-references GB:021123; NID:g758428; PID:g755821
GENETICS
#gene
#ena
##cross-references FlyBase:FBgn0000578
phosphoprotein
SUMMARY
#length 684 #molecular-weight 72049 #checksum 8892
Query Match 70.1%; Score 61; DB 2; Length 684;
Best Local Similarity 70.0%; Pred. No. 2.20e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 1; Gaps 1;
Db 79 QWRDSRFVYG 88
||||: |||
QY 283 QWRDTR-VYG 291

RESULT 3
ENTRY S22531 #type complete
TITLE pathogenesis-related protein 1b - common tobacco
ORGANISM #formal_name Nicotiana tabacum #common_name common tobacco
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
08-Sep-1997
ACCESSIONS S22531
REFERENCE S22531
#authors Eyal, Y.; Sagee, O.; Fluhr, R.
#journal Plant Mol. Biol. (1992) 19:589-599
#title Dark-induced accumulation of a basic pathogenesis-related
(PR-1) transcript and a light requirement for its induction
by ethylene.
#accession S22531
##molecule_type DNA
##residues 1-179 #label EYA
##cross-references EMBL:X66942; NID:g19969; PID:g19970
GENETICS
#gene prb-1b
CLASSIFICATION #superfamily pathogenesis-related leaf protein
SUMMARY #length 179 #molecular-weight 20120 #checksum 711
Query Match 65.5%; Score 57; DB 2; Length 179;
Best Local Similarity 60.0%; Pred. No. 1.02e+01;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 152 GNRGRRPYG 161
||||: |||
QY 282 GQWRDTRVYG 291

RESULT 4
ENTRY A35177 #type complete
TITLE chromate resistance protein B - Alcaligenes eutrophus plasmid
PMOL28
ORGANISM #formal_name Alcaligenes eutrophus
DATE 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change
09-Sep-1997
ACCESSIONS A35177
REFERENCE A35177
#authors Nies, A.; Nies, D.H.; Silver, S.
#journal J. Biol. Chem. (1990) 265:5648-5653
#title Nucleotide sequence and expression of a plasmid-encoded
chromate resistance determinant from Alcaligenes eutrophus.
#cross-references MUID:9020806
#accession A35177
#status preliminary
#molecule_type DNA
#residues 1-196 #label NIE
##cross-references GB:J05278; NID:g141911; PID:g141912
GENETICS
```

```
#genome
SUMMARY
#length 196 #molecular-weight 21697 #checksum 435
Query Match 65.5%; Score 57; DB 2; Length 196;
Best Local Similarity 55.6%; Pred. No. 1.02e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 188 QWRPASIVYG 196
||||: |||
QY 283 QWRDTRVYG 291

RESULT 5
ENTRY URXLA2 #type complete
TITLE peptidylglycine monooxygenase (EC 1.14.17.3) II precursor -
African clawed frog
ALTERNATE_NAMES C-terminal alpha-amidating enzyme II (AE-II); peptidyl
alpha-amidating enzyme II; peptidylglycine 2-hydroxylase II
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
13-Jun-1997
ACCESSIONS A27715
REFERENCE A27715
#authors Onsuaye, K.; Kitano, K.; Wada, Y.; Fuchimura, K.; Tanaka, S.;
Mizuno, K.; Matsuo, H.
#journal Biochem. Biophys. Res. Commun. (1988) 150:1275-1281
#title Cloning of cDNA encoding a new peptide C-terminal
alpha-amidating enzyme having a putative membrane-spanning
domain from Xenopus laevis skin.
#cross-references MUID:88134244
#accession A27715
##molecule_type mRNA
##residues 1-875 #label OHS
##cross-references GB:M20191
##experimental_source skin
COMMENT This copper protein is one of two enzymes that catalyze the
oxidation of peptidylglycine to the corresponding peptidyl
(2-hydroxyglycine) in the presence of ascorbate; the product is
unstable and dismutates to glyoxylate and the corresponding
desglycine peptide alpha-amide. The presence of such a C-terminal
alpha-amide structure is essential for the biological activity of
many peptide hormones. The other enzyme is peptidyl
alpha-amidating enzyme I.
CLASSIFICATION #superfamily peptidylglycine monooxygenase II;
peptidylglycine monooxygenase I homology
KEYWORDS copper; glycoprotein; monooxygenase; oxidoreductase;
transmembrane protein
FEATURE
1-24 #domain signal sequence #status predicted #label SIG\
24-39 #domain propeptide #status predicted #label PRO\
40-385 #product peptidylglycine monooxygenase II #status
predicted #label MAT\
133-344 #domain peptidylglycine monooxygenase I homology #label
PGM\
764-787 #domain transmembrane #status predicted #label TMN\
465,662,743 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 875 #molecular-weight 97084 #checksum 94
Query Match 65.5%; Score 57; DB 1; Length 875;
Best Local Similarity 55.6%; Pred. No. 1.02e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 788 RWRKVRMYG 796
||||: |||
QY 283 QWRDTRVYG 291

RESULT 6
ENTRY SI7855 #type complete
TITLE peptidylglycine monooxygenase (EC 1.14.17.3) - African clawed
frog
ALTERNATE_NAMES peptidylhydroxyglycine N-C lyase
```



```

ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change
08-Sep-1997
ACCESSIONS SI7855
REFERENCE Iwasaki, Y.; Kawahara, T.; Shimoi, H.; Suzuki, K.; Ghisalbal,
#authors O.; Kangawa, K.; Matsuo, H.; Nishikawa, Y.
#journal Eur. J. Biochem. (1991) 201:551-559
#title Purification and cDNA cloning of Xenopus laevis skin
peptidylhydroxylase N-C lyase, catalyzing the second
reaction of C-terminal alpha-amidation.
#cross-references MUID:92037609
#accession SI7855
#status preliminary
#molecule_type mRNA
#residues 1-935 #label IWA
#cross-references EMBL:X62771; NID:g64530; PID:g64531
CLASSIFICATION #superfamily peptidylglycine monooxygenase II;
peptidylglycine monooxygenase I homology
oxidoreductase
KEYWORDS #domain peptidylglycine monooxygenase I homology #label
FEATURE 131-342
PGM
SUMMARY #length 935 #molecular-weight 103379 #checksum 1873
Query Match 65.5%; Score 57; DB 2; Length 935;
Best Local Similarity 55.6%; Pred. No. 1.02e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 847 RWRKVRMYG 855
:|:|:|
QY 283 QWRDTRVYG 291

RESULT 7
ENTRY #type complete
TITLE hypothetical protein sl11468 - Synecocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S77365
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.;
Nakanura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yanada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S77365
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-312 #label KAN
#cross-references EMBL:D90806; GB:AB001339; NID:g1652492; PID:d1018201;
PID:g1652547
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
GENETICS
#start_codon GTG
SUMMARY #length 312 #molecular-weight 36151 #checksum 7413
Query Match 64.4%; Score 56; DB 2; Length 312;
Best Local Similarity 50.0%; Pred. No. 1.49e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 223 RWNARVY 230

```

```

QY 283 QWRDTRVY 290
:|:|:|:|
RESULT 8
ENTRY #type complete
TITLE protein phosphatase AB11 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 25-Oct-1994 #sequence_revision 01-Dec-1995 #text_change
09-Sep-1997
ACCESSIONS A54588
REFERENCE A54588
#authors Meyer, K.; Leube, M.P.; Grill, E.
#journal Science (1994) 264:1452-1455
#title A protein phosphatase 2C involved in ABA signal transduction
in Arabidopsis thaliana.
#cross-references MUID:94255767
#accession A54588
#status preliminary
#molecule_type DNA
#residues 1-434 #label MEY
#cross-references GB:X78886; NID:g509418; PID:g509419
GENETICS
#gene AB11
#introns 183/3; 280/3; 316/1
SUMMARY #length 434 #molecular-weight 47491 #checksum 4069
Query Match 64.4%; Score 56; DB 2; Length 434;
Best Local Similarity 55.6%; Pred. No. 1.49e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 299 QWNGARVFG 307
:|:|:|:|
QY 283 QWRDTRVYG 291

RESULT 9
ENTRY #type complete
TITLE hypothetical protein MJ1028 - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
21-Dec-1998
ACCESSIONS C64428
REFERENCE A64300
#authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton,
R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann,
J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Science (1996) 273:1058-1073
#title Complete genome sequence of the methanogenic archaeon,
Methanococcus jannaschii.
#cross-references MUID:96337999
#accession C64428
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-686 #label BUL
#cross-references GB:U67545; GB:L77117; NID:g1591680; PID:g1591683;
TIGR:MJ1028; PID:g1511054
GENETICS
#map_position FOR959556-961616
#start_codon TTG
CLASSIFICATION #superfamily Archaeoglobus fulgidus probable DNA
topoisomerase VI chain B
SUMMARY #length 686 #molecular-weight 78791 #checksum 9521

```

```
Query Match 64.4%; Score 56; DB 2; Length 686;
Best Local Similarity 70.0%; Pred. No. 1.49e+01;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 193 GKWRGTRVEG 202
    ||| ||| |
Qy 282 GQWRDTRVYG 291

RESULT 10
ENTRY C71305 #type complete
TITLE probable leucyl-tRNA synthetase (leuS) - syphilis spirochete
ORGANISM #formal_name Treponema pallidum subsp. pallidum #common_name
syphilis spirochete
DATE 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
17-Mar-1999
ACCESSIONS C71305
REFERENCE A71250
#authors Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.;
Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton,
R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod,
M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson,
D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald,
L.; Artlich, P.; Bowman, C.; Cotton, M.D.; Fujii, C.;
Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey,
L.; Weidman, J.; Smith, H.O.; Venter, J.C.
#journal Science (1998) 281:375-388
#title Complete genome sequence of Treponema pallidum, the syphilis
spirochete
#cross-references MUID:98332770
#accession C71305
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-878 #label COL
#cross-references GB:AE001234; GB:AE000520; NID:g3322881; PID:g3322882
#experimental_source strain Nichols
GENETICS
#gene TP0586
#classification superfamily leucine--tRNA ligase
#keywords aminocyl-tRNA synthetase; protein biosynthesis
SUMMARY #length 878 #molecular-weight 98990 #checksum 1126

Query Match 64.4%; Score 56; DB 2; Length 878;
Best Local Similarity 71.4%; Pred. No. 1.49e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 16 WRDKRVF 22
    ||| |||
Qy 284 WRDTRVY 290

RESULT 11
ENTRY H70544 #type complete
TITLE hypothetical protein Rv0525 - Mycobacterium tuberculosis
(strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS H70544
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellon, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
```

```
the complete genome sequence.
#cross-references MUID:98295987
#accession H70544
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-202 #label COL
#cross-references GB:295558; GB:AL123456; NID:g3261781; PID:e316853;
PID:g2113987
#experimental_source strain H37Rv
GENETICS
#gene Rv0525
SUMMARY #length 202 #molecular-weight 22212 #checksum 5815

Query Match 63.2%; Score 55; DB 2; Length 202;
Best Local Similarity 75.0%; Pred. No. 2.15e+01;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 100 GAWRDPRV 107
    ||| |||
Qy 282 GQWRDTRV 289

RESULT 12
ENTRY T00076 #type fragment
TITLE hypothetical protein KIAA0462 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
22-Jan-1999
ACCESSIONS T00076
REFERENCE Z14085
#authors Seki, N.; Ohira, M.; Nagase, T.; Ishikawa, K.; Miyajima, N.;
Nakajima, D.; Nomura, N.; Ohara, O.
#journal DNA Res. (1997) 4:345-349
#title Characterization of cDNA clones in size-fractionated cDNA
libraries from human brain.
#accession T00076
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-2276 #label SEK
#cross-references EMBL:AB007931; NID:dl225340; PID:dl033269
#experimental_source brain
GENETICS
#map_position 1
#note KIAA0462
SUMMARY #length 2276 #checksum 7906

Query Match 63.2%; Score 55; DB 2; Length 2276;
Best Local Similarity 85.7%; Pred. No. 2.15e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 2143 QWRATRV 2149
    ||| |||
Qy 283 QWRDTRV 289

RESULT 13
ENTRY TVFF7L #type complete
TITLE kinase-related protein sevenless - fruit fly (Drosophila
melanogaster)
CONTAINS protein-tyrosine kinase (BC 2.7.1.112)
ORGANISM #formal_name Drosophila melanogaster
DATE 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change
05-Sep-1997
ACCESSIONS A28912; A50107; A28827
REFERENCE A28912
#authors Basler, K.; Hafen, E.
#journal Cell (1988) 54:299-311
#title Control of photoreceptor cell fate by the sevenless protein
requires a functional tyrosine kinase domain.
#cross-references MUID:88282538
#accession A28912
#molecule_type DNA
```

```

##residues      1-2554  ##label BAS
##cross-references GB:J03158; NID:g158418; PID:g158419
REFERENCE
#authors      Hafen, E.; Basler, K.; Edstroem, J.E.; Rubin, G.M.
#journal      Science (1987) 236:55-63
#title        Sevenless, a cell-specific homeotic gene of Drosophila,
              encodes a putative transmembrane receptor with a tyrosine
              kinase domain.
#cross-references MUID:87177965
#accession    A60107
#status       nucleic acid sequence not shown
#molecule_type DNA
##residues      2062-2554  ##label HAF
REFERENCE
#authors      Bowtell, D.D.L.; Simon, M.A.; Rubin, G.M.
#journal      Genes Dev. (1988) 2:620-634
#title        Nucleotide sequence and structure of the sevenless gene of
              Drosophila melanogaster.
#cross-references MUID:88329706
#accession    A28827
#molecule_type DNA; mRNA
##residues      'RSSAS',1-391,'V',393-1822,'Q',1824-2270,'C',2272-2554
              ##label BOW
##cross-references GB:X13665
#note         392-Met, 1000-Leu, 1364-Val, 1668-Val, 1703-His,
              1730-Lys, 1731-Glu, 1741-Met, and 2271-Arg were also
              found
COMMENT       The sevenless gene controls the development of the R7 class of
              photoreceptor cells.
GENETICS
#gene         sev
#map_position X10A1-A2
#introns      227/3; 252/3; 274/3; 406/3; 486/3; 596/2; 2150/1; 2242/3;
              2307/3; 2456/2
CLASSIFICATION
#superfamily sevenless; fibronectin type III repeat homology;
              LDL receptor WYTD-containing repeat homology; protein
              kinase homology
KEYWORDS      ATP; autophosphorylation; glycoprotein; phosphoprotein;
              phosphotransferase; photoreceptor; receptor; transmembrane
              protein; tyrosine-specific protein kinase
FEATURE
102-122       #domain transmembrane #status predicted #label TMN1\
437-528       #domain fibronectin type III repeat homology #label FN3\
2124-2147     #domain transmembrane #status predicted #label TMN2\
2207-2488     #domain protein kinase homology #label KIN\
2215-2223     #region protein kinase ATP-binding motif\
129,481,505,617,
647,966,1228,1313,
1353,1550,1557,
1639,1725,1756,
1804,1889,1947,
2073
2242          #binding_site carbohydrate (Asn) (covalent) #status
              predicted\
SUMMARY       #active_site Lys #status predicted
              #length 2554 #molecule-weight 287108 #checksum 3353
Query Match   63.2%; Score 55; DB 1; Length 2554;
Best Local Similarity 62.5%; Pred. No. 2.15e+01;
Matches       5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db           1006 DNRGRVY 1013
              :|||
Qy           283 QWRDTRVY 290
              :|||
RESULT       14
ENTRY        ERBP4Z
TITLE        gene 4 protein - phage PZA
ORGANISM     #formal_name phage PZA
#note        host Bacillus subtilis
DATE         30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change
              17-Jul-1998

##residues      1-125  ##label PAC
##cross-references GB:M11813; GB:M13904; GB:M13905; NID:g216046;
              PID:g216049
GENETICS
#gene         4
#classification #superfamily phage PZA gene 4 protein
#keywords       early protein; transcription regulation
#summary        #length 125 #molecule-weight 15090 #checksum 647
Query Match   62.1%; Score 54; DB 1; Length 125;
Best Local Similarity 50.0%; Pred. No. 3.10e+01;
Matches       4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db           90 WREHVHYA 97
              :||:
Qy           284 WRDTRVYG 291
              :||:

RESULT       15
ENTRY        ERBP49
TITLE        early protein gp4 - phage phi-29
ALTERNATE_NAMES transcription activator p4
ORGANISM     #formal_name phage phi-29
#note        host Bacillus subtilis
DATE         15-Nov-1984 #sequence_revision 15-Nov-1984 #text_change
              05-Feb-1999
ACCESSIONS  C93439; A04284; S09124
REFERENCE     A93439
#authors      Escarmis, C.; Salas, M.
#journal      Nucleic Acids Res. (1982) 10:5785-5798
#title        Nucleotide sequence of the early genes 3 and 4 of
              bacteriophage psi29.
#cross-references MUID:83064518
#accession    C93439
#molecule_type DNA
##residues      1-125  ##label ESC
#note         the DNA sequence of ochre-suppressible mutant sus4(56)
              contains the termination codon TAA rather than the CAA
              codon at the position corresponding to Gln-42 of the
              wild type
REFERENCE     A91493
#authors      Yoshikawa, H.; Ito, J.
#journal      Gene (1982) 17:323-335
#title        Nucleotide sequence of the major early region of
              bacteriophage phi29.
#cross-references MUID:82262795
#accession    A04284
#status       translation not shown
#molecule_type DNA
##residues      1-125  ##label YOS
#cross-references GB:V01155; GB:J02478; NID:g15659; PID:g15669
#note         the authors assign the gene 4 protein to a different
              open reading frame; the sequence shown corresponds to
              ORF 9 of the authors; with the exception of the
              initiator Met, the reported composition based on the
              translated sequence agrees with that of the sequence
              shown
REFERENCE     S09124
#authors      Rojo, F.; Zaballos, A.; Salas, M.
#journal      J. Mol. Biol. (1990) 211:713-725
#title        Bend induced by the phage phi29 transcriptional activator in
              the viral late promoter is required for activation.
#cross-references MUID:90189140
#accession    S09124

```

```

##status      not compared with conceptual translation
##molecule_type DNA
##residues    110-125 ##label ROJ

GENETICS
#gene         4
#map_position 22-20
#description  early protein gp4 is believed to be a positive regulator of
               late transcription; it may function as a sigma-like
               component of the Bacillus RNA polymerase
CLASSIFICATION #superfamily phage P2A gene 4 protein
KEYWORDS       early protein; sigma factor; transcription
SUMMARY        #length 125 #molecular-weight 15133 #checksum 457

Query Match      62.1%; Score 54; DB 1; Length 125;
Best Local Similarity 50.0%; Pred. No. 3.10e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 90 WREVVHVA 97
   ||: ||:
QY 284 WRDTRVYG 291

```

Search completed: Thu Jul 8 18:45:31 1999
Job time : 16 secs.

MPERCH_PP protein - protein database search, using Smith-Waterman algorithm

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Thu Jul 8 18:44:12 1999; MasPar time 3.38 Seconds
Tabular output not generated. 83.514 Million cell updates/sec

Title: >US-09-041-236-2
Description: (282-291) from US09041236.pep (19 of 45)
Perfect Score: 87
Sequence: 1 GQWRDTRVYG 10

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28258293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 26.551; Variance 36.788; scale 0.722

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	61	70.1	391	1	PCL_ECTHA COENZYME A LIGASE-LIKE	6.82e-01
2	59	67.8	284	1	TRBG_AGR6 CONJUGAL TRANSFER PROT	1.61e+00
3	57	65.5	196	1	CHRB_ALCEU CHRB PROTEIN	3.73e+00
4	57	65.5	875	1	AMD2_XENLA PEPTIDYL-GLYCINE ALPHA	3.73e+00
5	56	64.4	434	1	P2C1_ARATH PROTEIN PHOSPHATASE 2C	5.64e+00
6	56	64.4	686	1	TP6B_MEPJA TYPE II DNA TOPOISOMER	5.64e+00
7	56	64.4	878	1	SYL_TREPA LEUCYL-TRNA SYNTHETASE	5.64e+00
8	55	63.2	2554	1	7LES_DROME SEVENLESS PROTEIN (EC	8.48e+00
9	54	62.1	125	1	VG4_BPP2A LATE GENES ACTIVATOR (1.27e+01
10	54	62.1	125	1	VG4_BPP2 LATE GENES ACTIVATOR (1.27e+01
11	54	62.1	253	1	KANU_BACSP KANAMYCIN NUCLEOTIDYL	1.27e+01
12	54	62.1	253	1	KANU_STAAU KANAMYCIN NUCLEOTIDYL	1.27e+01
13	54	62.1	380	1	VASP_HUMAN VASODILATOR-STIMULATED	1.27e+01
14	54	62.1	384	1	VASP_CANFA VASODILATOR-STIMULATED	1.27e+01
15	53	60.9	163	1	PHZA_PSEFL PHENAZINE BIOSYNTHESIS	1.88e+01
16	53	60.9	187	1	Y4GA_RHISN HYPOTHETICAL 20.9 KD P	1.88e+01
17	53	60.9	192	1	Y4DX_RHISN HYPOTHETICAL 20.6 KD P	1.88e+01
18	53	60.9	494	1	CPS1_HUMAN CYTOCHROME P450 XXIB (1.88e+01
19	53	60.9	767	1	RRPO_RCNMV PUTATIVE RNA-DIRECTED	1.88e+01
20	53	60.9	1091	1	SVI_TREPA ISOLEUCYL-TRNA SYNTHET	1.88e+01
21	53	60.9	1432	1	SK13_YEAST SUPERKILLER 3 PROTEIN	1.88e+01
22	53	60.9	1432	1	PTPW_MOUSE PROTEIN-TYROSINE PHOSP	1.88e+01
23	53	60.9	1452	1	PTPM_HUMAN PROTEIN-TYROSINE PHOSP	1.88e+01

24	53	60.9	3412	1	POLG_TBVS	GENOME POLYPROTEIN [CO	1.88e+01
25	53	60.9	3414	1	POLG_LANVT	GENOME POLYPROTEIN [CO	1.88e+01
26	53	60.9	3430	1	POLG_WNV	GENOME POLYPROTEIN [CO	1.88e+01
27	53	60.9	3433	1	POLG_KUNJM	GENOME POLYPROTEIN [CO	1.88e+01
28	52	59.8	213	1	I141_HUMAN	IMMUNOGLOBULIN-RELATED	2.78e+01
29	52	59.8	256	1	YN80_YEAST	HYPOTHETICAL 28.6 KD P	2.78e+01
30	52	59.8	328	1	YDDP_ECOLI	HYPOTHETICAL ABC TRANS	2.78e+01
31	52	59.8	372	1	KC21_YEAST	CASEIN KINASE II, ALPHA	2.78e+01
32	52	59.8	419	1	FAAA_HUMAN	FUMARYLACETOACETASE (E	2.78e+01
33	52	59.8	457	1	GAC4_CHICK	GAMMA-AMINOBUTYRIC-ACI	2.78e+01
34	52	59.8	461	1	TNR1_RAT	TUMOR NECROSIS FACTOR	2.78e+01
35	52	59.8	496	1	CPS1_BOVIN	CYTOCHROME P450 XXIA1	2.78e+01
36	52	59.8	497	1	CATA_MICLU	CATALASE (EC 1.11.1.6)	2.78e+01
37	52	59.8	538	1	YIEC_ECOLI	POTENTIAL OUTER MEMBRA	2.78e+01
38	52	59.8	928	1	NRP_XENLA	NEUROFILIN PRECURSOR (2.78e+01
39	51	58.6	146	1	GPR7_MOUSE	PROBABLE G PROTEIN-COU	4.09e+01
40	51	58.6	328	1	GPR7_HUMAN	PROBABLE G PROTEIN-COU	4.09e+01
41	51	58.6	366	1	Y433_ARCFU	HYPOTHETICAL PROTEIN A	4.09e+01
42	51	58.6	409	1	YADE_ECOLI	HYPOTHETICAL 46.3 KD P	4.09e+01
43	51	58.6	512	1	Y00K_MCTU	HYPOTHETICAL 53.6 KD P	4.09e+01
44	51	58.6	648	1	Y095_TREPA	HYPOTHETICAL PROTEIN T	4.09e+01
45	51	58.6	3432	1	POLG_JAEVJ	GENOME POLYPROTEIN [CO	4.09e+01

ALIGNMENTS

ALIGNMENTS

RESULT 1	ID	PCL_ECTHA	STANDARD;	PRT;	391 AA.
AC	P42516;				
DT	01-NOV-1995 (REL. 32, CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	COENZYME A LIGASE-LIKE PROTEIN (FRAGMENT).				
GN	ECL.				
OS	ECTOTHIORHODOSPIRA HALOPHILA.				
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ECTOTHIORHODOSPIRACEAE;				
OC	HALORHODOSPIRA.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BN9626;				
RX	MEDLINE; 96283619.				
RA	KORT R., HOFF W.D., VAN WEST M., KROON A.R., HOFFER S.M.,				
RA	VLIIG K.H., CRIELAARD W., VAN BEUWEN J.J., HELLINGWERF K.J.;				
RT	"The xanthopsins: a new family of eubacterial blue-light				
RT	photoreceptors.";				
RL	EMBO J. 15:3209-3218(1996).				
RN	[2]				
RP	SEQUENCE OF 1-112 FROM N.A.				
RC	STRAIN-BN9626;				
RX	MEDLINE; 95072006.				
RA	BACA M., BORGSTAHL G.E., BOISSINOT M., BURKE P.M., WILLIAMS D.R.,				
RA	SLATER K.A., GETZOFF E.D.;				
RT	"Complete chemical structure of photoactive yellow protein: novel				
RT	thioester-linked 4-hydroxycinnamyl chromophore and photocycle				
RT	chemistry.";				
RL	BIOCHEMISTRY 33:14369-14377(1994).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch)				
CC	-----				
CC	EMBL; X98897; E251788; -				
DR	EMBL; 017017; G602430; -				
FW	LIGASE.				
KT	NON_TER 391 391				
SQ	SEQUENCE 391 AA; 42438 MW; F1A7E739 CRC32;				

Query Match 70.1%; Score 61; DB 1; Length 391;
Best Local Similarity 75.0%; Pred. No. 6.82e-01;

```
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 284 QWRDERVF 291
   |||||
Qy 283 QWRDTRVY 290

RESULT 2
ID TRBG_AGR6 STANDARD; PRT; 284 AA.
AC P54915;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CONJUGAL TRANSFER PROTEIN TRBG PRECURSOR.
GN TRBG.
OS AGROBACTERIUM TUMEFACIENS.
OG PLASMID PTIAGNC.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96312368.
RA ALT-MORBE J., STRIKER J.L., FLOUQUA C., LI P.L., FARRAND S.K.,
RA WINANS S.C.;
RT "The conjugal transfer system of Agrobacterium tumefaciens
RT octopine-type Ti plasmids is closely related to the transfer system
RT of an IncP plasmid and distantly related to Ti plasmid vir genes.";
RL J. BACTERIOL. 178:4248-4257(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J05278; G141912; -
DR PIR; A35177; A35177.
KW PLASMID; MEMBRANE; CHROMATE RESISTANCE.
SQ SEQUENCE 196 AA; 21697 MW; 12AFD5E5 CRC32;

Query Match 65.5%; Score 57; DB 1; Length 196;
Best Local Similarity 55.6%; Pred. No. 3.73e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 188 QWRPASIVG 196
   ||| : |||
Qy 283 QWRDTRVYG 291

RESULT 4
ID AMD2_XENLA STANDARD; PRT; 875 AA.
AC P12890;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PEPTIDYL-GLYCINE ALPHA-AMIDATING MONOOXYGENASE II PRECURSOR
DE (EC 1.14.17.3) (PEPTIDE C-TERMINAL ALPHA-AMIDATING ENZYME II) (AE-II).
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKIN;
RX MEDLINE; 88134244.
RA OHSUYE K., KITANO K., WADA Y., FUCHIMURA K., TANAKA S., MIZUNO K.,
RA MATSUO H.;
RT "Cloning of cDNA encoding a new peptide C-terminal alpha-amidating
RT enzyme having a putative membrane-spanning domain from Xenopus laevis
RT skin.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 150:1275-1281(1988).
CC -!- FUNCTION: C-TERMINAL ALPHA-AMIDATING OF BIOLOGICAL PEPTIDES.
CC -!- CATALYTIC ACTIVITY: PEPTIDYLGLYCINE + ASCORBATE + O(2) -
CC PEPTIDYL(2-HYDROXYGLYCINE) + DEHYDROASCORBATE + H(2)O.
CC (THE PRODUCT IS UNSTABLE AND DISMUTATES TO GLYOXYLATE AND THE
CC CORRESPONDING DESGLYCINE PEPTIDE AMIDE).
CC -!- COFACTOR: COPPER, AND ASCORBATE.
CC -!- SIMILARITY: BELONGS TO THE COPPER TYPE II, ASCORBATE-DEPENDENT
CC MONOOXYGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M19032; G214015; -
DR PIR; A27715; URXLA2.
DR PROSITE; PS00084; CU2_MONOOXYGENASE_1; 1.
DR PROSITE; PS00085; CU2_MONOOXYGENASE_2; 1.
DR PFAM; PF01082; Cu2_monooxygen; 1.
KW OXIDOREDUCTASE; MONOOXYGENASE; COPPER; VITAMIN C; TRANSMEMBRANE;
KW GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 39 POTENTIAL.
```

```

FT CHAIN 40 875 PEPTIDYL-GLYCINE ALPHA-AMIDATING
FT MONOOXYGENASE II
FT INTRAGRANULAR (POTENTIAL).
FT TRANSSEM 764 787
FT DOMAIN 40 763
FT TRANSSEM 764 787
FT DOMAIN 788 875
FT DOMAIN 358 364
FT DOMAIN 387 390
FT DOMAIN 852 856
FT CARBOHYD 465 465
FT CARBOHYD 562 662
FT CARBOHYD 743 743
SQ SEQUENCE 875 AA; 97084 MW; 06EEF456 CRC32;

Query Match 65.5%; Score 57; DB 1; Length 875;
Best Local Similarity 55.6%; Pred. No. 3.73e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 788 RWRKVRMYG 796
||| |||
QY 283 QWRDTRVYG 291

RESULT 5
ID P2C1ARATH STANDARD; PRT; 434 AA.
AC P49597;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROTEIN PHOSPHATASE 2C AB11 (EC 3.1.3.16) (PP2C).
GN AB11.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPALTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, LANDSBERG ERECTA;
RX MEDLINE: 94255767.
RA MEYER K., LEUBE M.P., GRILL E.;
RT "A protein phosphatase 2C involved in ABA signal transduction in
RT Arabidopsis thaliana."
RL SCIENCE 264:1452-1455(1994).
CC -1- FUNCTION: INVOLVED IN ABSCISIC ACID (ABA) SIGNALLING PATHWAY.
CC -1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O -> A PROTEIN +
CC ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
CC -1- COFACTOR: BINDS TWO MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PP2C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78886; G509419;
CC PROSITE; PS00018; EF_HAND; 1.
CC DR PROSITE; PS01032; PP2C; 1.
CC PFAM; PF00481; PP2C; 2.
CC HSP; P35813; I460
CC HYDROLASE; MAGNESIUM; MANGANESE; MULTIGENE FAMILY; CALCIUM-BINDING.
KW CA_BIND 93 104
FT DOMAIN 417 420 POLY-VAL.
FT VARIANT 180 180 G -> D (WILTY PHENOTYPE AND ABA-
FT INSENSITIVE SEED GERMINATION AND GROWTH).
SQ SEQUENCE 434 AA; 47491 MW; 3E76B56E CRC32;

Query Match 64.4%; Score 56; DB 1; Length 434;
Best Local Similarity 55.6%; Pred. No. 5.64e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 299 QWNGARVFG 307

```

```

QY 283 QWRDTRVYG 291
||| |||
RESULT 6
ID TP6B_METJA STANDARD; PRT; 686 AA.
AC Q58434;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYPE II DNA TOPOISOMERASE VI SUBUNIT B (EC 5.99.1.3).
GN TOP6B OR MJ1028.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
OC METHANOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE: 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA OVERLAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
RA SCOTT J.L., GEORHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HORST M.A., KAINE B.P., BORODOVSKY M.C.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL SCIENCE 273:1058-1073(1996).
CC -1- FUNCTION: RELAXES BOTH POSITIVE AND NEGATIVE SUPERURNS AND
CC EXHIBITS A STRONG DECATENASE ACTIVITY. THE B SUBUNIT BINDS ATP (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SUBUNIT: HETEROTETRAMER OF TWO SUBUNITS A AND TWO SUBUNITS B (BY
CC SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67545; G1591683;
CC TIGR; MJ1028;
CC ISOMERASE; TOPOISOMERASE; DNA-BINDING; ATP-BINDING.
KW ISOMERASE; TOPOISOMERASE; DNA-BINDING; ATP-BINDING.
SQ SEQUENCE 686 AA; 78791 MW; A7D73B0B CRC32;

Query Match 64.4%; Score 56; DB 1; Length 686;
Best Local Similarity 70.0%; Pred. No. 5.64e+00;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 193 GKWRGTRVYG 202
||| |||
QY 282 QWRDTRVYG 291

RESULT 7
ID SYL_TREPA STANDARD; PRT; 878 AA.
AC O83595;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE LEUCYL-TRNA SYNTHETASE (EC 6.1.1.4) (LEUCINE--TRNA LIGASE) (LEURS).
GN LEUS OR TP0586.
OS TREPONEMA PALLIDUM.
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; TREPONEMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;

```

```
RX MEDLINE; 98332770.
RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA SODERGEN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA McDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete.";
RL SCIENCE 281:375-388(1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-LEUCINE + TRNA(LEU) = AMP +
CC -!- SUBSTRATE: L-LEUCYL-TRNA(LEU).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001234; G3322882; -.
CC TIGR; TP0586; -.
CC PROSITE; PS00178; AA-TRNA LIGASE I; 1.
CC AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
CC FT SIMILAR 45 "HIGH" REGION.
CC FT SIMILAR 634 638 "KMSKS" REGION.
CC FT BINDING 637 637 ATP (BY SIMILARITY).
CC FT SEQUENCE 878 AA; 98991 MW; 78D81CF0 CRC32;
CC -----
Query Match 64.4%; Score 56; DB 1; Length 878;
Best Local Similarity 71.4%; Pred. No. 5.64e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 16 WRDKRVF 22
QY 284 WRDTRVY 290
III III;
RESULT 8
ID 7LES_DROME STANDARD; PRT; 2554 AA.
AC P13368;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SEVENLESS PROTEIN (EC 2.7.1.112).
GN SEV.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN CANTON-S;
RX MEDLINE; 88282538.
RA BASLER K., HAFEN E.;
RT "Control of photoreceptor cell fate by the sevenless protein requires
RT a functional tyrosine kinase domain.";
RL CELL 54:299-311(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN OREGON-R;
RX MEDLINE; 88329706.
RA BOWTELL D.L.L., SIMON M.A., RUBIN G.M.;
RT "Nucleotide sequence and structure of the sevenless gene of
RT Drosophila melanogaster.";
RL GENES DEV. 2:620-634(1988).
RN [3]
RP IDENTIFICATION OF FN-III REPEATS.
```

```
RX MEDLINE; 90199889.
RA NORTON P.A., HYNES R.O., RESS D.J.G.;
RT "Sevenless: seven found?";
RL CELL 61:15-16(1990).
CC -!- FUNCTION: RECEPTOR FOR AN EXTRACELLULAR SIGNAL REQUIRED TO
CC INSTRUCT A CELL TO DIFFERENTIATE INTO A R7 PHOTORECEPTOR. THE
CC LIGAND FOR SEV IS THE BOSS (BRIDE OF SEVENLESS) PROTEIN ON THE
CC SURFACE OF THE NEIGHBORING R8 CELL.
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE -- ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBUNIT: MAY FORM A COMPLEX WITH DRK AND SOS.
CC -!- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
CC PROTEIN KINASES. SEVENLESS SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 7 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- CAUTION: UNCLEAR WHETHER THE POTENTIAL MEMBRANE SPANNING REGION
CC NEAR THE N-TERMINUS IS PRESENT AS A TRANSMEMBRANE DOMAIN IN THE
CC NATIVE PROTEIN OR SERVES AS A CLEAVED SIGNAL SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13666; G8579; ALT_INIT.
CC EMBL; J03158; G158419; -.
CC PIR; A28912; TVF7L.
CC FLYBASE; FBgn0003366; sev.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PFAM; PF00041; fn3; 6;
CC PFAM; PF00069; pkinase; 1.
CC HSSP; P11362; IFGI.
CC TRANSFERASE; TYROSINE-PROTEIN KINASE; TRANSMEMBRANE; ATP-BINDING;
KW PHOSPHORYLATION; RECEPTOR; VISION; REPEAT.
FT DOMAIN 1 2123 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 102 122 POTENTIAL.
FT TRANSMEM 2124 2147 POTENTIAL.
FT DOMAIN 2148 2554 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 311 431 FIBRONECTIN TYPE-III.
FT DOMAIN 436 528 FIBRONECTIN TYPE-III.
FT DOMAIN 822 921 FIBRONECTIN TYPE-III.
FT DOMAIN 1298 1392 FIBRONECTIN TYPE-III.
FT DOMAIN 1680 1794 FIBRONECTIN TYPE-III.
FT DOMAIN 1797 1897 FIBRONECTIN TYPE-III.
FT DOMAIN 1898 1988 FIBRONECTIN TYPE-III.
FT DOMAIN 2038 2046 POLY-ARG.
FT DOMAIN 2209 2485 PROTEIN KINASE.
FT NP_BIND 2215 2223 ATP (BY SIMILARITY).
FT BINDING 2242 2242 ATP (BY SIMILARITY).
FT MUTAGEN 2242 2242 K->M: INACTIVATES THE PROTEIN.
FT MOD_RES 2380 2380 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 30 30 POTENTIAL.
FT CARBOHYD 129 129 POTENTIAL.
FT CARBOHYD 481 481 POTENTIAL.
FT CARBOHYD 505 505 POTENTIAL.
FT CARBOHYD 617 617 POTENTIAL.
FT CARBOHYD 647 647 POTENTIAL.
FT CARBOHYD 966 966 POTENTIAL.
FT CARBOHYD 1228 1228 POTENTIAL.
FT CARBOHYD 1313 1313 POTENTIAL.
FT CARBOHYD 1353 1353 POTENTIAL.
FT CARBOHYD 1550 1550 POTENTIAL.
FT CARBOHYD 1557 1557 POTENTIAL.
FT CARBOHYD 1639 1639 POTENTIAL.
FT CARBOHYD 1725 1725 POTENTIAL.
FT CARBOHYD 1756 1756 POTENTIAL.
FT CARBOHYD 1804 1804 POTENTIAL.
FT CARBOHYD 1889 1889 POTENTIAL.
```


FT CARBOHYD 1947 1947 POTENTIAL.
FT CARBOHYD 2073 2073 POTENTIAL.
FT VARIANT 392 392 M -> V.
FT VARIANT 1668 1668 A -> V.
FT VARIANT 1703 1703 N -> H.
FT VARIANT 1730 1730 R -> K.
FT VARIANT 1731 1731 G -> E.
FT VARIANT 1741 1741 V -> M.
FT VARIANT 2271 2271 R -> C.
FT CONFLICT 1823 1823 E -> Q (IN REF. 2).
SQ SEQUENCE 2554 AA; 287107 MW; 1143D891 CRC32;

Query Match 63.2%; Score 55; DB 1; Length 2554;

Best Local Similarity 62.5%; Pred. No. 8.48e+00;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1006 DNRGRVY 1013

QY 283 QWRDTRVY 290

RESULT 9
ID VG4_BPP2A STANDARD; PRT; 125 AA.
AC P06952;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE LATE GENES ACTIVATOR (EARLY PROTEIN GP4).
GN 4.

OS BACTERIOPHAGE PZA.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; PODOVIRIDAE.
RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 86056991.

RA PACES V., VLCEK C., URBANEK P., HOSTOMSKY Z.;

RT "Nucleotide sequence of the major early region of Bacillus subtilis

phage PZA, a close relative of phi 29.";

RL GENE 38:45-56(1985).

CC -!- FUNCTION: THIS PROTEIN IS BELIEVED TO BE A POSITIVE REGULATOR OF

LATE TRANSCRIPTION. IT MAY FUNCTION AS A SIGMA-LIKE COMPONENT OF

THE HOST RNA POLYMERASE.

CC -!- SIMILARITY: BELONGS TO THE GP4 PODOVIRUSES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M11813; G216049; -.

DR PIR; F24528; ERBP4Z.

KW EARLY PROTEIN: TRANSCRIPTION REGULATION; SIGMA FACTOR;

KW DNA-DIRECTED RNA POLYMERASE.

FT DNA_BIND 77 96 BY SIMILARITY.

SQ SEQUENCE 125 AA; 15090 MW; B8AE03E1 CRC32;

Query Match 62.1%; Score 54; DB 1; Length 125;

Best Local Similarity 50.0%; Pred. No. 1.27e+01;

Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 90 WREHVHYA 97

QY 284 WRDTRVYG 291

RESULT 10
ID VG4_BPPH2 STANDARD; PRT; 125 AA.
AC P03682;

DT 21-JUL-1986 (REL. 01, CREATED)

DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE LATE GENES ACTIVATOR (EARLY PROTEIN GP4).
GN 4.
OS BACTERIOPHAGE PHI-29.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; PODOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 83064518.
RA ESCARMIS C., SALAS M.;

"Nucleotide sequence of the early genes 3 and 4 of bacteriophage phi 29.";

RL NUCLEIC ACIDS RES. 10:5785-5798(1982).

RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 82262795.
RA YOSHIKAWA H., ITO J.;

RT "Nucleotide sequence of the major early region of bacteriophage phi 29.";

RL GENE 17:323-335(1982).

CC -!- FUNCTION: THIS PROTEIN IS BELIEVED TO BE A POSITIVE REGULATOR OF
CC LATE TRANSCRIPTION. IT MAY FUNCTION AS A SIGMA-LIKE COMPONENT OF
CC THE HOST RNA POLYMERASE.

CC -!- SIMILARITY: BELONGS TO THE GP4 PODOVIRUSES FAMILY.

CC -!- CAUTION: REF.2 CONFIRMS THE DNA SEQUENCE. THESE AUTHORS ASSIGNED
CC THE GENE 4 PROTEIN TO A DIFFERENT OPEN READING FRAME (ORF); THE
CC SEQUENCE SHOWN CORRESPONDS TO ORF 9 OF THE AUTHORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; V01155; G15669; -.

DR PIR; A04284; ERBP49.

KW EARLY PROTEIN: TRANSCRIPTION REGULATION; SIGMA FACTOR;

KW DNA-DIRECTED RNA POLYMERASE.

FT DNA_BIND 77 96 BY SIMILARITY.

SQ SEQUENCE 125 AA; 15133 MW; 892361B7 CRC32;

Query Match 62.1%; Score 54; DB 1; Length 125;

Best Local Similarity 50.0%; Pred. No. 1.27e+01;

Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 90 WREHVHYA 97

QY 284 WRDTRVYG 291

RESULT 11
ID KANU_BACSP STANDARD; PRT; 253 AA.
AC P05058;

DT 13-AUG-1987 (REL. 05, CREATED)

DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)

DT 01-SEP-1995 (REL. 31, LAST ANNOTATION UPDATE)

DE KANAMYCIN NUCLEOTIDYLTRANSFERASE (EC 2.7.7.-).

OS KNT OR KAN.

GN BACILLUS SP.

OG PLASMID PRBH1, AND PLASMID PTB913.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;

OC BACILLUS.

RN [1]

RP SEQUENCE FROM N.A.

RC PLASMID-PRBH1.

RX MEDLINE; 86174350.

RA MULLER R.E., ANO T., IMANAKA T., AIBA S.;

RT "Complete nucleotide sequences of Bacillus plasmids pUB110dB, pRBH1

and its copy mutants.";

RL MOL. GEN. GENET. 202:169-171(1986).

RN [2]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-7.

RC PLASMID-PTB913;

RX MEDLINE: 85006820.
RA MATSUMURA M., KATAKURA Y., IMANAKA T., AIBA S.:
RT "Enzymatic and nucleotide sequence studies of a
RT kanamycin-inactivating enzyme encoded by a plasmid
RT bacilli in comparison with that encoded by plasmid
RL J. BACTERIOL. 160:413-420(1984).
RN [3]
RP SEQUENCE FROM N.A.
RC PLASMID-PTB913;
RX MEDLINE: 90016790.
RA VAN DER LELIE D., BRON S., VENEMA G., OSKAM L.:
RT "Similarity of minus origins of replication and flanking
RT frames of plasmids pub110, PTB913 and pMV158".
RL NUCLEIC ACIDS RES. 17:7283-7294(1989).
CC -!- FUNCTION: INACTIVATES THE ANTIBIOTIC KANAMYCIN BY
CC TRANSFER OF A NUCLEOTIDYL GROUP FROM NUCLEOSIDE
CC AS ATP TO THE 4'-HYDROXYL GROUP OF THE AMINOGLYCOSIDE.
CC -!- SUBUNIT: HOMODIMER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a
CC between the Swiss Institute of Bioinformatics and the
CC the European Bioinformatics Institute. There are no
CC use by non-profit institutions as long as its content is
CC modified and this statement is not removed. Usage by
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K02551; G294383; -
DR EMBL: X03409; G580911; -
DR EMBL: X15670; G40080; ALT_INIT.
DR HSP: P05057; LKNY.
KW ANTIBIOTIC RESISTANCE; TRANSFERASE; PLASMID.
SQ SEQUENCE 253 AA; 28824 MW; 2BA2c4C1 CRC32;

Query Match 62.1%; Score 54; DB 1; Length 253;
Best Local Similarity 60.0%; Pred. No. 1.27e+01;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 148 GKNRIRVQG 157
QY 282 GQWRDTRVYG 291

RESULT 12
ID KANU_STAAU STANDARD; PRT; 253 AA.
AC P05057;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE KANAMYCIN NUCLEOTIDYLTRANSFERASE (EC 2.7.7.-) (NEO(R)).
GN KNT OR KAN.
OS STAPHYLOCOCCUS AUREUS.
OG PLASMID PUB110.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC STAPHYLOCOCCUS.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86174350.
RA MULLER R.E., ANO T., IMANAKA T., AIBA S.:
RT "Complete nucleotide sequences of Bacillus plasmids
RT bacilli in comparison with that encoded by a plasmid
RL J. BACTERIOL. 160:413-420(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86301853.
SQ SEQUENCE 253 AA; 28797 MW; 87BD4F15 CRC32;

RA BASHKIROV V.I., MIL'SHINA N.V., PROZOROV A.A.:
RT "Nucleotide sequence and physical map of kanamycin-resistant plasmid
RT pub110 from Staphylococcus aureus".
RL GENETIKA 22:1081-1092(1986).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86206447.
RA MCKENZIE T., HOSHINO T., TANAKA T., SUBOKA N.:
RT "The nucleotide sequence of PUB110; some salient features in relation
RT to replication and its regulation".
RL PLASMID 15:93-103(1986).
RN [5]
RP REVISIONS.
RX MEDLINE: 87204663.
RA MCKENZIE T., HOSHINO T., TANAKA T., SUBOKA N.:
RT "Correction. A revision of the nucleotide sequence and functional map
RT of PUB110".
RL PLASMID 17:83-84(1987).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE: 94032329.
RA SAKON J., LIAO H.H., KANIKULA A.M., BENNING M.M., RAYMENT I.,
RA HOLDEN H.M.:
RT "Molecular structure of kanamycin nucleotidyltransferase determined
RT to 3.0-A resolution".
RL BIOCHEMISTRY 32:11977-11984(1993).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE: 96027442.
RA PEDERSEN L.C., BENNING M.M., HOLDEN H.M.:
RT "Structural investigation of the antibiotic and ATP-binding sites in
RT kanamycin nucleotidyltransferase".
RL BIOCHEMISTRY 34:13305-13311(1995).
CC -!- FUNCTION: INACTIVATES THE ANTIBIOTIC KANAMYCIN BY CATALYZING THE
CC TRANSFER OF A NUCLEOTIDYL GROUP FROM NUCLEOSIDE TRIPHOSPHATES SUCH
CC AS ATP TO THE 4'-HYDROXYL GROUP OF THE AMINOGLYCOSIDE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- THE ENZYME ENCODED BY PTB913 IS MORE THERMOSTABLE THAN THAT
CC ENCODED BY PUB110.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03408; G581544; -
DR EMBL: M37273; G150535; -
DR EMBL: K02552; G294074; -
DR EMBL: M19465; G455307; -
DR PIR: A24456; A24456.
DR PDB: LKAN; 31-AUG-94.
DR PDB: LKNY; 17-AUG-96.
KW ANTIBIOTIC RESISTANCE; TRANSFERASE; PLASMID; 3D-STRUCTURE.
SQ SEQUENCE 253 AA; 28797 MW; 87BD4F15 CRC32;

Query Match 62.1%; Score 54; DB 1; Length 253;
Best Local Similarity 60.0%; Pred. No. 1.27e+01;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 148 GKNRIRVQG 157
QY 282 GQWRDTRVYG 291

RESULT 13
ID VASP_HUMAN STANDARD; PRT; 380 AA.
AC P50552;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE VASODILATOR-STIMULATED PHOSPHOPROTEIN (VASP).
GN VASP.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 95129547.
RA HAFNER C., JARCHAU T., REINHARD M., HOPPE J., LOHMANN S.M.,
WALTER U.;
RT "Molecular cloning, structural analysis and functional expression of
the proline-rich focal adhesion and microfilament-associated protein
VASP";
RL EMBO J. 14:19-27(1995).
RN [2]
RN SEQUENCE OF 151-160; 235-244 AND 267-282, AND PHOSPHORYLATION SITES.
RX MEDLINE: 94237860.
RA BUTT E., ABEL K., KRIEGER M., PALM D., HOPPE V., HOPPE J., WALTER U.;
RT "cAMP- and CGMP-dependent protein kinase phosphorylation sites of the
focal adhesion vasodilator-stimulated phosphoprotein (VASP) in vitro
and in intact human platelets";
RL J. BIOL. CHEM. 269:14509-14517(1994).
RN [3]
RN BINDING TO PROFILIN.
RX MEDLINE: 95255215.
RA REINHARD M., GIEHL K., ABEL K., HAFNER C., JARCHAU T., HOPPE V.,
JOCKUSH B.M., WALTER U.;
RT "The proline-rich focal adhesion and microfilament protein VASP is a
ligand for profilins";
RL EMBO J. 14:1583-1589(1995).
RN [4]
RN FUNCTION: ACTIN- AND PROFILIN-BINDING MICROFILAMENT-ASSOCIATED
PROTEIN. MAY ACT IN CONCERT WITH PROFILIN TO CONVEY SIGNAL
TRANSDUCTION TO ACTIN FILAMENT PRODUCTION.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: FOCAL ADHESIONS.
CC -1- PTM: MAJOR SUBSTRATE FOR CAMP-DEPENDENT (CAPK) AND CGMP-DEPENDENT
PROTEIN KINASE (CGPK) IN PLATELETS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 246389; G624964; -
CC MTM: 601703; -
CC PFAM: PF00568; WHI: 1.
CC PHOSPHORYLATION: REPEAT; ACTIN-BINDING.
FT MOD_RES 157 157 PHOSPHORYLATION (BY CAPK AND CGPK).
FT MOD_RES 239 239 PHOSPHORYLATION (BY CAPK AND CGPK).
FT MOD_RES 278 278 PHOSPHORYLATION (BY CAPK AND CGPK).
FT DOMAIN 118 122 POLY-PRO.
FT DOMAIN 170 186 POLY-PRO.
FT DOMAIN 215 222 POLY-GLY.
FT DOMAIN 259 262 POLY-GLY.
FT DOMAIN 322 325 POLY-SER.
SQ SEQUENCE 380 AA; 39830 MW; E44E1AF2 CRC32;

Query Match 62.1%; Score 54; DB 1; Length 380;
Best Local Similarity 70.0%; Pred. No. 1.27e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Db 81 QWRDAROVWG 90
QY 283 QWRDTR-VYG 291
||||:| | |

RESULT 14
ID VASP-CANFA STANDARD; PRT; 384 AA.
AC P50551;
DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE VASODILATOR-STIMULATED PHOSPHOPROTEIN (VASP).
GN VASP.
OS CANIS FAMILIARIS (DOG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
RN [1]
RN SEQUENCE FROM N.A.
RX TISSUE-KIDNEY;
RX MEDLINE: 95129547.
RA HAFNER C., JARCHAU T., REINHARD M., HOPPE J., LOHMANN S.M.,
WALTER U.;
RT "Molecular cloning, structural analysis and functional expression of
the proline-rich focal adhesion and microfilament-associated protein
VASP";
RL EMBO J. 14:19-27(1995).
RN [2]
RN FUNCTION: ACTIN- AND PROFILIN-BINDING MICROFILAMENT-ASSOCIATED
PROTEIN. MAY ACT IN CONCERT WITH PROFILIN TO CONVEY SIGNAL
TRANSDUCTION TO ACTIN FILAMENT PRODUCTION.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: FOCAL ADHESIONS.
CC -1- PTM: MAJOR SUBSTRATE FOR CAMP-DEPENDENT (CAPK) AND CGMP-DEPENDENT
PROTEIN KINASE (CGPK) IN PLATELETS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 246388; G624250; -
CC PFAM: PF00568; WHI: 1.
CC PHOSPHORYLATION: REPEAT; ACTIN-BINDING.
FT DOMAIN 118 124 POLY-PRO.
FT DOMAIN 173 189 POLY-PRO.
FT MOD_RES 160 160 PHOSPHORYLATION (BY CAPK AND CGPK).
FT MOD_RES 242 242 PHOSPHORYLATION (BY CAPK AND CGPK).
FT MOD_RES 281 281 PHOSPHORYLATION (BY CAPK AND CGPK).
FT MOD_RES 384 AA; 40413 MW; 1B236BE7 CRC32;
SQ SEQUENCE 384 AA; 40413 MW; 1B236BE7 CRC32;

Query Match 62.1%; Score 54; DB 1; Length 384;
Best Local Similarity 70.0%; Pred. No. 1.27e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Db 81 QWRDAROVWG 90
QY 283 QWRDTR-VYG 291
||||:| | |

RESULT 15
ID PHZA_PSEFL STANDARD; PRT; 163 AA.
AC O51787;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PHENAZINE BIOSYNTHESIS PROTEIN PHZA.
GN PHZA.
OS PSEUDOMONAS FLUORESCENS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
PSEUDOMONAS.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=NRRL B-15132 / 2-79;
RA MAVRODI D.V., KSENZENKO V.N., BORONIN A.M., COOK J.R.,
RA THOMASHOW L.S.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF THE ANTIBIOTIC,
CC PHENAZINE, A NITROGEN-CONTAINING HETEROCYCLIC MOLECULE HAVING
CC IMPORTANT ROLES IN VIRULENCE, COMPETITION AND BIOLOGICAL CONTROL.
CC -!- SIMILARITY: HIGH, TO P.FLUORESCENS PHZB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----

DR EMBL; L48616; G1045013; -.
KW ANTIBIOTIC BIOSYNTHESIS.
SQ SEQUENCE 163 AA; 18706 MW; 600DAC8F CRC32;

Query Match 60.9%; Score 53; DB 1; Length 163;
Best Local Similarity 50.0%; Pred. No. 1.88e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 82 QWHNVRVF 89

||: ||:

QY 283 QWRDTRVY 290

Search completed: Thu Jul 8 18:44:21 1999
Job time : 9 secs.

RX MEDLINE: 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL: AF030698; G3523115; -.
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 87; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 7.07e-06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 314 GOWRDRVYG 323
| | | | |
Qy 282 GOWRDRVYG 291

RESULT 3
ID O08719 PRELIMINARY; PRT; 393 AA.
AC O08719;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RNB6.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR;
RX MEDLINE: 97415794.
RA OHTA S., MINETA T., KIMOTO M., TABUCHI K.;
RT "Differential display cloning of a novel rat cDNA (RNB6) that shows
RT high expression in the neonatal brain revealed a member of Ena/VASP
RT family."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 237:307-312(1997).
DR EMBL: U70211; G2058462; -.
DR PFAM: PF00568; WH1; 1.
SQ SEQUENCE 393 AA; 42095 MW; 3B485732 CRC32;

Query Match 75.9%; Score 66; DB 11; Length 393;
Best Local Similarity 80.0%; Pred. No. 1.69e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 80 QWRDARQVYG 89
| | | | |
Qy 283 QWRDTR-VYG 291

RESULT 4
ID P70429 PRELIMINARY; PRT; 393 AA.
AC P70429;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENA-VASODILATOR STIMULATED PHOSPHOPROTEIN (ENA-VASP LIKE PROTEIN).
GN EVL.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97015079.
RA GERTLER F.B., NIEBUHR K., REINHARD M., WEHLAND J., SORIANO P.;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in
RT the control of microfilament dynamics."
RL CELL 87:227-239(1996).
DR EMBL: U72519; G1644453; -.
DR MGD; MGI:1194884; EVL.
DR PFAM: PF00568; WH1; 1.
SQ SEQUENCE 393 AA; 42123 MW; 80FA013A CRC32;

Query Match 75.9%; Score 66; DB 11; Length 393;
Best Local Similarity 80.0%; Pred. No. 1.69e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 80 QWRDARQVYG 89
| | | | |
Qy 283 QWRDTR-VYG 291

RESULT 5
ID P70430 PRELIMINARY; PRT; 541 AA.
AC P70430;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENABLED HOMOLOG (MENA PROTEIN).
GN ENAH OR MENA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97015079.
RA GERTLER F.B., NIEBUHR K., REINHARD M., WEHLAND J., SORIANO P.;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in
RT the control of microfilament dynamics."
RL CELL 87:227-239(1996).
DR EMBL: U72520; G1644455; -.
DR MGD; MGI:108360; ENAH.
DR PFAM: PF00568; WH1; 1.
SQ SEQUENCE 541 AA; 59642 MW; DA5430E9 CRC32;

Query Match 75.9%; Score 66; DB 11; Length 541;
Best Local Similarity 80.0%; Pred. No. 1.69e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 79 QWRDARQVYG 88
| | | | |
Qy 283 QWRDTR-VYG 291

RESULT 6
ID O93263 PRELIMINARY; PRT; 550 AA.
AC O93263;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE AVENA.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ITO M., OHASHI K.;
RT "Proteins related to Ena and Mena in chicken muscle tissues."
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AB017437; D1033982; -.
SQ SEQUENCE 550 AA; 60992 MW; 4D640816 CRC32;

Query Match 75.9%; Score 66; DB 13; Length 550;
Best Local Similarity 80.0%; Pred. No. 1.69e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 79 QWRDARQVYG 88
| | | | |
Qy 283 QWRDTR-VYG 291

RESULT 7
ID P70431 PRELIMINARY; PRT; 783 AA.
AC P70431;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)

```
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENABLED HOMOLOG (NEURAL VARIANT MENA+ PROTEIN).
GN ENAH OR MENA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SC SCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 97015079.
RA GERTLER F.B., NIEBUHR K., REINHARD M., WEHLAND J., SORIANO P.;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in
the control of microfilament dynamics.";
RL CELL 87:227-239(1996).
DR EMBL; U72521; G1644457; -.
DR MGD; MGI:108360; ENAH.
DR PFAM; PF00568; WHI; 1.
SQ SEQUENCE 783 AA; 83775 MW; 6E60B8C5 CRC32;

Query Match 75.9%; Score 66; DB 11; Length 783;
Best Local Similarity 80.0%; Pred. No. 1.69e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 79 QWRDARQVYG 88
||||:| |||
Qy 283 QWRDTR-VYG 291

RESULT 8
ID P70432 PRELIMINARY; PRT; 787 AA.
AC P70432;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENABLED HOMOLOG (NEURAL VARIANT MENA++ PROTEIN).
GN ENAH OR MENA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SC SCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 97015079.
RA GERTLER F.B., NIEBUHR K., REINHARD M., WEHLAND J., SORIANO P.;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in
the control of microfilament dynamics.";
RL CELL 87:227-239(1996).
DR EMBL; U72522; G1644459; -.
DR MGD; MGI:108360; ENAH.
DR PFAM; PF00568; WHI; 1.
SQ SEQUENCE 787 AA; 84298 MW; B5079049 CRC32;

Query Match 75.9%; Score 66; DB 11; Length 787;
Best Local Similarity 80.0%; Pred. No. 1.69e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 79 QWRDARQVYG 88
||||:| |||
Qy 283 QWRDTR-VYG 291

RESULT 9
ID P70433 PRELIMINARY; PRT; 802 AA.
AC P70433;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENABLED HOMOLOG (NEURAL VARIANT MENA+++ PROTEIN).
GN ENAH OR MENA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SC SCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 97015079.
RA GERTLER F.B., NIEBUHR K., REINHARD M., WEHLAND J., SORIANO P.;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in
the control of microfilament dynamics.";
RL CELL 87:227-239(1996).
DR EMBL; U72523; G1644461; -.
DR MGD; MGI:108360; ENAH.
DR PFAM; PF00568; WHI; 1.
SQ SEQUENCE 802 AA; 85844 MW; BDC99AE8 CRC32;

Query Match 75.9%; Score 66; DB 11; Length 802;
Best Local Similarity 80.0%; Pred. No. 1.69e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 79 QWRDARQVYG 88
||||:| |||
Qy 283 QWRDTR-VYG 291

RESULT 10
ID Q24035 PRELIMINARY; PRT; 684 AA.
AC Q24035;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENA POLYPEPTIDE.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95212903.
RA GERTLER F.B., CONER A.R., JUANG J.L., AHERN S.M., CLARK M.J.,
LIEBL E.C., HOFFMANN F.M.;
RT "enabled, a dosage-sensitive suppressor of mutations in the
Drosophila Abl tyrosine kinase, encodes an Abl substrate with SH3
domain-binding properties.";
RL GENES DEV. 9:521-533(1995).
DR EMBL; U21123; G755821; -.
DR FLYBASE; FBgn0000578; enb.
DR PFAM; PF00568; WHI; 1.
SQ SEQUENCE 684 AA; 72049 MW; 1F36F629 CRC32;

Query Match 70.1%; Score 61; DB 5; Length 684;
Best Local Similarity 70.0%; Pred. No. 1.51e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

Db 79 QWRDSKFVYG 88
||||:| |||
Qy 283 QWRDTR-VYG 291

RESULT 11
ID O66177 PRELIMINARY; PRT; 284 AA.
AC O66177;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TREG.
GN TREG.
OS AGROBACTERIUM TUMEFACIENS.
OG PLASMID PTI-SAKURA.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MAFF301001;
RX MEDLINE; 98193120.
RA SUZUKI K., OHTA N., HATTORI Y., URAJI M., KATO A., YOSHIDA K.;
RT "Novel structural difference between nopaline- and octopine-type trbJ
```

RT genes: construction of genetic and physical map and sequencing of
trb/trai and rep gene clusters of a new Ti plasmid pTi-SAKURA.";
RL BIOCHIM. BIOPHYS. ACTA 1396:1-7(1998).
DR EMBL: AB006858; D1026210; .
KW PLASMID.
SQ SEQUENCE 284 AA; 30594 MW; AD871FF2 CRC32;

Query Match 67.8%; Score 59; DB 2; Length 284;
Best Local Similarity 75.0%; Pred. No. 3.52e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 207 WRPTRVVS 214
|||
Qy 284 WRDTRVYG 291

RESULT 12 PRELIMINARY; PRT; 309 AA.
ID O43495
AC O43495;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE G PROTEIN-COUPLED RECEPTOR.
GN GPR35.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98140132.
RA O'DOWD B.F., NGUYEN T., MARCHESE A., CHENG R., LYNCH K.R.,
RA HENG H.H.Q., KOLAKOWSKI L.F. JR., GEORGE S.R.;
RT "Discovery of three novel G-protein-coupled receptor genes.";
RL GENOMICS 47:310-313(1998).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF027957; G2739109; .
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN.
SQ SEQUENCE 309 AA; 34226 MW; 8A9C4D52 CRC32;

Query Match 67.8%; Score 59; DB 4; Length 309;
Best Local Similarity 62.5%; Pred. No. 3.52e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 51 QWTETRIY 58
|||
Qy 283 QWRDTRVY 290

RESULT 13 PRELIMINARY; PRT; 511 AA.
ID O81709
AC O81709; 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PROTEIN PHOSPHATASE 2C.
GN P2C-HA.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LANDSBERG ERECTA;
RA RODRIGUEZ P.L.;
RT "Molecular cloning of a new protein phosphatase 2C (PP2C) with
homology to AB11 and AB12.";
RL PLANT MOL. BIOL. 0:0-0(0).
DR EMBL: AJ003119; E1300921; .
DR PROSITE; PS01032; PP2C; 1.
SQ SEQUENCE 511 AA; 55713 MW; CD780A2E CRC32;

Query Match 66.7%; Score 58; DB 10; Length 511;

Best Local Similarity 55.6%; Pred. No. 5.35e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 384 QWOGARVFG 392
|||
Qy 283 QWRDTRVYG 291

RESULT 14 PRELIMINARY; PRT; 783 AA.
ID O13956
AC O13956;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 90.1 KD PROTEIN C23H4.15 IN CHROMOSOME I.
GN SPAC23H4.15.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA BROWN D., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: TO YEAST YDL060W.
DR EMBL: Z98977; E339926; .
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 142 162 POTENTIAL.
FT TRANSMEM 643 663 POTENTIAL.
FT DOMAIN 451 454 POLY-GLU.
SQ SEQUENCE 783 AA; 90125 MW; 4592049D CRC32;

Query Match 66.7%; Score 58; DB 3; Length 783;
Best Local Similarity 62.5%; Pred. No. 5.35e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 246 HWRDSRSY 253
|||
Qy 283 QWRDTRVY 290

RESULT 15 PRELIMINARY; PRT; 179 AA.
ID O04106
AC O04106;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PATHOGENESIS-RELATED PROTEIN 1B PRECURSOR (PR-1B).
GN PRB-1B.
OS NICOTIANA TABACUM (COMMON TOBACCO).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; NICOTIANA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, SAMSUN;
RX MEDLINE; 92329719.
RA EYAL Y., SAGEE O., FLUHR R.;
RT "Dark-induced accumulation of a basic pathogenesis-related (PR-1)
transcript and a light requirement for its induction by ethylene.";
RL PLANT MOL. BIOL. 19:589-599(1992).
DR EMBL: X65942; G19970; .
DR PIR; S22531; S22531.
DR FRAM; PF00188; SCP; 1.
KW PATHOGENESIS-RELATED PROTEIN; MULTIGENE FAMILY.
SQ SEQUENCE 179 AA; 20120 MW; 7812F870 CRC32;

Query Match 65.5%; Score 57; DB 10; Length 179;
Best Local Similarity 60.0%; Pred. No. 8.07e+00;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 152 GNRGQRPYG 161

QY 282 GQWRDTRVIG 291
| | | | |

Search completed: Thu Jul 8 18:44:55 1999
Job time : 12 secs.

THIS PAGE BLANK (USPTO)

W P E R L H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:48:09 1999; Maspar time 11.06 Seconds
36.524 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (335-353) from US09041236.pep (20 of 45)
Perfect Score: 136
Sequence: 1 LPDQOPIPTFTFOVADRRP 19

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-genes35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 19.409; Variance 64.639; scale 0.300

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	59	43.4	745	2	R11809 RNA dependant RNA pol	5.50e+01	
2	56	41.2	260	19	W00637 ILTV ORF10 product.	1.07e+02	
3	56	41.2	260	21	W06789 ILTV unique short reg	1.07e+02	
4	56	41.2	572	3	R20105 ERK3.	1.07e+02	
5	54	39.7	357	32	W60327 Bacillus thuringiensis	1.66e+02	
6	54	39.7	669	16	R86408 Human matrix metallopro	1.66e+02	
7	54	39.7	825	2	R11254 Human IL-4 receptor.	1.66e+02	
8	54	39.7	887	25	W19856 Rat semaphorin Z.	1.66e+02	
9	54	39.7	2351	3	F50319 Human antihemophilic	1.66e+02	
10	53	39.0	171	16	R84656 STLVpan-p rex protein	2.06e+02	
11	53	39.0	240	22	W15273 Salmonella tyrosine p	2.06e+02	
12	53	39.0	266	12	R62863 Korean-type Hepatitis	2.06e+02	
13	53	39.0	1274	39	W89253 Human ALP.	2.06e+02	
14	53	39.0	3010	6	R30816 Polypeptide coded by	2.06e+02	
15	52	38.2	348	38	W81969 Mouse E124 protein #2	2.55e+02	
16	52	38.2	357	32	W60226 Bacillus thuringiensis	2.55e+02	

17	52	38.2	416	19	W06795	High Osmolality Glyce	2.55e+02
18	52	38.2	440	1	P94878	Bik 1 incorporated wi	2.55e+02
19	52	38.2	440	3	R14907	BIK1 protein.	2.55e+02
20	52	38.2	441	13	R71381	Vaccinia virus semaph	2.55e+02
21	52	38.2	529	34	W60878	Rhizoctonia solani (I	2.55e+02
22	52	38.2	529	22	W16301	Rhizoctonia solani la	2.55e+02
23	52	38.2	529	13	R72328	Laccase RSLac3.	2.55e+02
24	52	38.2	529	36	W76310	Rhizoctonia solani (I	2.55e+02
25	51	37.5	212	3	R13898	Nitrite hydratase bet	3.16e+02
26	51	37.5	261	3	R13505	P. denitrificans COB K	3.16e+02
27	51	37.5	284	10	R50071	NANBH virus antigenic	3.16e+02
28	51	37.5	924	6	R35445	Human TPO lacking ami	3.16e+02
29	51	37.5	1062	26	W22721	Antigenic protein der	3.16e+02
30	50	36.8	142	6	R28985	Thyroid N-acetyl-glyc	3.91e+02
31	50	36.8	411	20	W09300	Cuphea class II acyl-	3.91e+02
32	50	36.8	411	18	W02081	Cuphea C14:0-ACP thio	3.91e+02
33	50	36.8	412	7	R38309	Sequence of the P450-	3.91e+02
34	50	36.8	443	27	W32110	Human extracellular/e	3.91e+02
35	50	36.8	464	24	W18094	Xenopus alpha-signal	3.91e+02
36	50	36.8	763	30	W41734	Human TRAF-2 kinase.	3.91e+02
37	50	36.8	777	39	W85601	Hexosaminidase enzyme	3.91e+02
38	50	36.8	782	32	W57436	Teramyl-linker-CBD f	3.91e+02
39	50	36.8	1027	22	W17057	Candida albicans chit	3.91e+02
40	50	36.8	1664	34	W43106	C. thermocellum OlpB	3.91e+02
41	50	36.8	1728	3	R13144	Deleted in Colorectal	3.91e+02
42	49	36.0	299	13	R68549	Hepatitis C virus (HC	4.82e+02
43	49	36.0	1038	38	W86246	Human BMP receptor ki	4.82e+02
44	49	36.0	1038	38	W86247	Mouse BMP receptor ki	4.82e+02
45	49	36.0	2436	1	P92050	Sequence encoded in t	4.82e+02

ALIGNMENTS

RESULT 1
ID R11809 standard; Protein; 745 AA.
AC R11809;
DT 27-JUN-1991 (first entry)
DE RNA dependant RNA polymerase (replicase).
KW RNA dependant RNA polymerase; replicase; plant viruses.
OS Tumour necrosis virus
PN EP-425004-A.
PD 02-MAY-1991.
PF 03-OCT-1990; 202627.
PR 03-OCT-1989; NL-002452.
PR 27-JUL-1990; NL-001711.
PA (AVEV-) Aveve NV.
PA (CLOV-) Clovis Matton NV.
PA (SOLV-) Solvay and Cie.
PI Van Haute E, Ameloot P, De Lafonteyne J, Fiers W;
DR WPI; 91-126522/18.
DR N-PSDB; Q11641.
PT Recombinant DNA comprising RNA derived sequences - used for
PT protecting organisms, esp. plants from virus infection or for
PT producing protein or RNA
PS Disclosure; fig 4; 30pp; English.
CC The corresp. tumour necrosis virus (TNV) replicase DNA sequence is
CC co-transfected, into a host cell genome, with an expression cass-
CC ette comprising RNA virus derived sequences. Pref. the cassette
CC comprises (a) two 12-1250 bp inverted LTR sequences and, between
CC these, (b) at least one sequence, from an RNA virus dependant on
CC replicase, having cis elements for replication and no coding info.
CC for replicase or viral coat protein. The transformed host is
CC subsequently protected against RNA viruses dependant on the repli-
CC case for replication. By incorporating non-viral sequences the
CC expression cassette can be used to produce, in an inducible- or
CC tissue specific manner, protein(s)/peptide(s) or RNA(s), eg
CC ribozymes, antisense RNAs or double-stranded RNAs.
SQ Sequence 745 AA;

Query Match 43.4%; Score 59; DB 2; Length 745;
Best Local Similarity 58.3%; Pred. No. 5.50e+01;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 722 lptdyqvsvgeh 733
|||:|:|:|
QY 341 IPTETQVADRH 352

RESULT 2
ID W00637 standard; Protein: 260 AA.
AC W00637;
DT 19-NOV-1996 (first entry)
DE ILTV ORF10 product.
KW Infectious laryngotracheitis virus; ILTV; herpesvirus;
KW attenuation; vector; vaccine; chicken; poultry; immunisation.
OS Infectious laryngotracheitis virus.
FH Key Location/Qualifiers
FT misc_difference 113
FT /note= "unidentified amino acid"
PN W09508622-A1.
PD 30-MAR-1995.
PF 16-SEP-1994; U10628.
PR 24-SEP-1993; US-126597.
PA (SYTR) SYNTRO CORP.
PI Cochran MD, Wild MA;
DR WPI: 95-139591/18.
DR N-PSDB; T33504.
PT Recombinant attenuated infectious laryngotracheitis virus - for use
PT in vaccines to protect poultry from infection from the virus, also
PT methods of distinguishing between vaccinated and naturally infected
PT birds
PS Example 1; Page 105; 177pp; English.
CC Open reading frame 10 (ORF10), spanning nucleotides 12,665-13,447
CC of the unique short region (T33504) of infectious laryngotracheitis
CC virus (ILTV), codes for a 27,898 protein (W00637) of unknown
CC function.
SQ Sequence 260 AA;

Query Match 41.2%; Score 56; DB 19; Length 260;
Best Local Similarity 47.1%; Pred. No. 1.07e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 125 lpsdqaapartyssaqr 141
||| | : | : | : |
QY 335 LPDQQPIPTETQVADR 351

RESULT 3
ID W06789 standard; Protein: 260 AA.
AC W06789;
DT 02-JUN-1997 (first entry)
DE ILTV unique short region ORF10 polypeptide product.
KW ILTV; vaccine; vector; attenuation; poultry;
KW avian infectious bronchitis virus; Newcastle disease virus;
KW infectious bursal disease virus of chickens;
KW Marek's disease virus; herpesvirus.
OS Infectious laryngotracheitis virus USDA strain 8302.
FH Key Location/Qualifiers
FT misc_difference 112
FT /note= "unidentified amino acid"
PN W09629396-A1.
PD 26-SEP-1996.
PF 21-MAR-1996; U03916.
PR 23-MAR-1995; US-410121.
PR 06-JUN-1995; US-468190.
PA (SYTR) SYNTRO CORP.
PI Cochran MD, Wild MA;
DR WPI: 96-443172/44.
DR N-PSDB; T44384;
DR N-PSDB; T44385.
PT Recombinant infectious laryngotracheitis virus with deletion in the
PT glycoprotein G, q1 or u52 gene, etc. - useful for vaccines against
PT infectious laryngotracheitis in poultry
PS Example 11; Page 113-114; 216pp; English.
CC A 260-amino acid polypeptide (W06789) of unknown function is
CC encoded by open reading frame 10 of the unique short region

CC (T44384) of infectious laryngotracheitis virus (ILTV). Attenuated
CC recombinant ILTVs have at least one deletion in the unique short
CC region and can be used in vaccines against ILTV, and also in
CC multivalent vaccines for one or more avian diseases.
SQ Sequence 260 AA;

Query Match 41.2%; Score 56; DB 21; Length 260;
Best Local Similarity 47.1%; Pred. No. 1.07e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 125 lpsdqaapartyssaqr 141
||| | : | : | : |
QY 335 LPDQQPIPTETQVADR 351

RESULT 4
ID R20105 standard; Protein: 572 AA.
AC R20105;
DT 31-MAR-1992 (first entry)
DE ERK3.
KW Microtubule-associated protein 2; central nervous system;
KW extracellular signal-regulated kinase; ERK; CNS; MAP2;
KW nerve growth factor; insulin; neutrophin; FUS3; KSS1.
OS Rat rattus.
FH Key Location/Qualifiers
FT misc_difference 56
FT /note= "conserved among all protein kinases"
FT misc_difference 58
FT /note= "conserved among all protein kinases"
FT misc_difference 61
FT /note= "conserved among all protein kinases"
FT misc_difference 63
FT /note= "conserved among all protein kinases"
FT misc_difference 76
FT /note= "conserved among all protein kinases"
FT misc_difference 78
FT /note= "conserved among all protein kinases"
FT misc_difference 94
FT /note= "conserved among all protein kinases"
FT misc_difference 181
FT /note= "conserved among all protein kinases"
FT misc_difference 185
FT /note= "conserved among all protein kinases"
FT misc_difference 200
FT /note= "conserved among all protein kinases"
FT misc_difference 201
FT /note= "conserved among all protein kinases"
FT misc_difference 202
FT /note= "conserved among all protein kinases"
FT misc_difference 228
FT /note= "conserved among all protein kinases"
FT misc_difference 229
FT /note= "conserved among all protein kinases"
FT misc_difference 230
FT /note= "conserved among all protein kinases"
FT misc_difference 243
FT /note= "conserved among all protein kinases"
FT misc_difference 248
FT /note= "conserved among all protein kinases"
FT misc_difference 333
FT /note= "conserved among all protein kinases"
PN W09119008-A.
PD 12-DEC-1991.
PD 03-JUN-1991; U03894.
PR 01-JUN-1990; US-532004.
PR 16-MAY-1991; US-701544.
PA (REG-) REGENERON PHARM INC.
PA (TEXA) UNIV OF TEXAS SYSTEM.
PI Boulton TG, Cobb MH, Yancopoulos GD, Nye S, Panayotatos N;
DR WPI: 92-007489/01.
DR N-PSDB; Q20282.
PT DNA encoding MAP2 kinase enzyme and vectors or host cells - are
PT for assaying cellular factor (e.g. NGF), and drug screening

PS Disclosure; Fig 3B; 9pp; English.
 CC The indicated amino acids are residues most conserved among all
 CC protein kinases. Cells producing the kinase can be used for
 CC assaying cellular factor activity where the activation of the kinase
 CC serves as an indicator of activity. Cpd. detected include those
 CC having nerve growth factor-like activity, insulin-like activity or
 CC neurotrophin mols.
 CC ERK1-3, represented in Q0260-62, have a mol. wt. of 41-48 kD
 CC (SDS-PAGE) and have an N-terminal portion which shows homology with
 CC yeast FUS3 or KRS1 protein kinases.
 SQ Sequence 572 AA;

Query Match 41.2%; Score 56; DB 3; Length 572;
 Best Local Similarity 31.3%; Pred. No. 1.07e+02;
 Matches 5; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Db 350 fptdeptsfshfied 365
 :|::|::|::|
 QY 335 LPDQPIPTETFOVAD 350

RESULT 5

ID W60227 standard; Protein; 357 AA.

AC W60227;
 DT 28-SEP-1998 (first entry)
 DE Bacillus thuringiensis MIS-6 insecticidal toxin 196F3.
 KW insecticide; pesticide; toxin; MIS-6; delta-endotoxin;
 KW biological control; lepidopteran; coleopteran.
 OS Bacillus thuringiensis strain PS196F3 (NRRL B-21872).
 FH Key Location/Qualifiers
 FT Misc_difference 3
 FT Misc_difference 6 /note= "encoded by CNT"
 FT Misc_difference 32 /note= "encoded by TAY"
 FT Misc_difference 32 /note= "encoded by GAM"
 FT Misc_difference 76 /note= "encoded by NAT"
 FT Misc_difference 98 /note= "encoded by AMA"
 FT Misc_difference 108 /note= "encoded by KAA"
 FT Misc_difference 147 /note= "encoded by GNT"
 PN W09818932-A2.
 PD 07-MAY-1998.
 PF 30-OCT-1997; U19804.
 PR 30-OCT-1996; US-029848.
 PA (MYCO) MYCOGEN CORP.
 PI Dullum CJ, Feitelson JS, Loewer D, Muller-Cohn J,
 PI Narva KE, Schmeits JL, Schnepf HE, Schwab G, Stamp L,
 PI Stockhoff BA;
 PI WPI; 98-272226/24.
 DR N-PSDB: V30310.
 DR Bacillus thuringiensis isolates - used for producing pesticidal
 PT toxins and nucleotide sequences for control of lepidopterans and
 PT coleopterans

PS Claim 5; Page 90-92; 139pp; English.
 CC This polypeptide comprises a novel soluble toxin of Bacillus
 CC thuringiensis (B.t.) strain PS196F3 (NRRL B-21872). The toxin
 CC belongs to the novel MIS-6 family of B.t. toxins that have toxicity
 CC against non-mammalian pests. Its amino acid sequence was deduced
 CC from a novel DNA fragment (see V30310) obtained by PCR from
 CC cellular genomic DNA of PS196F3. Disclosed and claimed are novel
 CC B.t. isolates and toxins (see W60218-32) that have activity against
 CC lepidopteran and/or coleopteran pests, isolated genes, probes
 CC and primers (see V30288-321 and T99734-87) useful for production
 CC of the toxins and for the identification and characterisation of
 CC these toxins, and transformed hosts, particularly plant and
 CC bacterial hosts. The invention provides 8 entirely new families of
 CC toxins, including MIS-6, from B.t. isolates. MIS toxins have the
 CC additional ability to form pores in cell membranes, and can be used
 CC to facilitate entry of a second agent into a target cell.

SQ Sequence 357 AA;

Query Match 39.7%; Score 54; DB 32; Length 357;
 Best Local Similarity 50.0%; Pred. No. 1.66e+02;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 130 epipdnafqld 141
 :|||::|::|
 QY 339 QPIPTETFOVAD 350

RESULT 6

ID R66408 standard; Protein; 669 AA.

AC R66408;
 DT 15-MAY-1996 (first entry)
 DE Human matrix metalloproteinase MMPm2.
 KW Human; matrix metalloproteinase; MMP; extracellular matrix; inhibitor;
 KW activator; zinc-binding region.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT region 109..115
 FT /note= "conserved region within pro-sequence,
 FT contains a Cys residue"
 FT region 128..133
 FT /note= "conserved region within pro-sequence"
 FT domain 259..269
 FT /label= catalytic
 FT /note= "contains 3 His residues"
 FT region 626..644
 FT /label= hydrophobic
 PN W09535171-A2.
 PD 21-SEP-1995.
 PF 17-MAR-1995; D00357.
 PR 17-MAR-1994; DE-409663.
 PR 21-OCT-1994; DE-438838.
 PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.
 PI Hinzmann B, Will H;
 PI WPI; 95-336975/43.
 DR N-PSDB: T03438.
 DR New matrix metalloproteinase(s) and DNA encoding them - also vectors,
 PT recombinant cells and complexes with ligands, useful as proteolytic
 PT agents and for identification of specific inhibitors and activators
 PT Claim 10; Page 31; 85pp; German.
 PS The present sequence is that of human matrix metalloproteinase
 CC MMPm2. The protein has mol. wt. 75813 and comprises a signal
 CC peptide, pro-region, conserved catalytic domain and haemopexin-
 CC homologous sequence, all typical of known metalloproteinases. MMPm2
 CC differs from known metalloproteinases in having a hydrophobic region
 CC near the C-terminus, suggesting that the protein is membrane-bound.
 CC MMP proteins can hydrolyse extracellular matrix proteins and are
 CC involved in certain pathologies. The new MMP will be useful for
 CC detecting MMP inhibitors or activators and to generate diagnostic
 CC antibodies.
 SQ Sequence 669 AA;

Query Match 39.7%; Score 54; DB 16; Length 669;
 Best Local Similarity 47.4%; Pred. No. 1.66e+02;
 Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Db 306 pdgqpqtqlptvtprp 324
 :|::|::|
 QY 336 PDQQPIPTETFO-VADHRP 353

RESULT 7

ID R11254 standard; Protein; 825 AA.

AC R11254;
 DT 30-MAY-1991 (first entry)
 DE Human IL-4 receptor.
 KW Interleukin-4; soluble; IgE.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT region 1..125

```

FT FT /label< signal sequence
FT FT 26..232
FT FT /label< extracellular domain
FT FT 233..256
FT FT /label< transmembrane region
FT FT 257..825
FT FT /label< cytoplasmic domain
PN EP-419091-A.
PD 27-MAR-1991.
PF 05-SEP-1990: 309700.
PR 07-SEP-1989: US-404179.
PR 20-MAR-1990: US-496449.
PA (SCHE ) SCHERING CORP.
PI Galizzi JP, Harada N, Miyajima A;
DR WPI: 91-088700/13.
PT Nucleic acid encoding mammalian interleukin-4 receptor - used as
PT antagonists of interleukin-4 in treating conditions associated
PT with excess Ige prodn. including allergic conditions.
PS Disclosure; Fig 5; 18pp; English.
CC The sequence was deduced from a cDNA clone isolated from a cDNA
CC library prep. from mRNA the human Burkitt lymphoma cell line, BL41
CC and the human multifactor dependent myeloid cell line, TF1. The
CC BL41 library was screened using mouse IL-4R cDNA (Q11055) and a
CC partial nucleotide sequence retrieved from a positive clone, 4A2 was
CC used to probe the TF1 library to isolate clone TF1-2.2. The sequence
CC was used to prepare modified genes encoding soluble IL-4 receptors.
CC See also R11255.
CC Sequence 825 AA;
SQ

Query Match 39.7%; Score 54; DB 2; Length 825;
Best Local Similarity 50.0%; Pred. No. 1.66e+02;
Matches 9; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

Db 645 papvptlefgl-drep 661
QY 336 PDQOPIPTETFOVADHRP 353

RESULT 8
ID W19856 standard; Protein; 887 AA.
AC W19856;
DE 13-FEB-1998 (first entry)
DE Rat semaphorin 2.
KW Semaphorin 2; central nerve extension; rat; human; inhibitor;
KW central nerve regeneration promoter.
OS Rattus norvegicus.
PN WQ9720928-A1.
PD 12-JUN-1997.
PF 02-DEC-1996; J03517.
PR 31-OCT-1996; JP-307205.
PR 06-DEC-1995; JP-345187.
PA (SUMU ) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 97-319775/29.
DR N-PSDB; T72107.
PT Semaphorin 2 and the gene encoding it - also inhibitors of its
PT action which can be used as promoters of central nerve regeneration
PS Claim 3; Page 66-71; 106pp; Japanese.
CC This sequence represents the rat semaphorin 2. This protein sequence, and
CC the human semaphorin 2 protein (see W19857) can be used for screening
CC possible candidates for activity as semaphorin 2 inhibitors. Proteins
CC containing all or part of the semaphorin 2 sequence are useful as central
CC nerve extension inhibitors. Semaphorin 2 inhibitors identified by
CC screening can be used as promoters of central nerve regeneration.
CC Sequence 887 AA;
SQ

Query Match 39.7%; Score 54; DB 25; Length 887;
Best Local Similarity 60.0%; Pred. No. 1.66e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 760 papvptptes 769
QY 336 PDQOPIPTET 345

```

```

RESULT 9
ID P50319 standard; Protein; 2351 AA.
AC P50319;
DE 23-NOV-1991 (first entry)
DE Human antihaemophilic factor.
KW Human antihaemophilic factor; haemophilia; factor VIIIC.
OS Homo sapiens.
PN W08501961-A.
PD 09-MAY-1985.
PF 12-OCT-1984; U01641.
PR 28-OCT-1983; US-546650.
PR 24-OCT-1984; US-644036.
PA (GENE-) Genetics Institute.
PI Toole JJ Jr;
DR WPI: 85-122479/20.
DR N-PSDB; N50375.
PT Prodn. of human and porcine factor VIIIC - by using recombinant DNA
PT techniques for their cellular prodn.
PS Claim 41; page 54; 71pp; English.
CC Human antihaemophilic factor is useful for the treatment of haemophilia.
CC Sequence 2351 AA;
SQ

Query Match 39.7%; Score 54; DB 3; Length 2351;
Best Local Similarity 42.1%; Pred. No. 1.66e+02;
Matches 8; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

Db 750 lsknnpieptfsqsrhp 768
QY 335 LPDQOPIPTETFOVADHRP 353

RESULT 10
ID R84656 standard; Protein; 171 AA.
AC R84656;
DE 26-MAY-1996 (first entry)
DE STLvpan-p rex protein.
KW STLvpan-p; simian T-cell lymphotropic virus; animal model;
KW probe; primer; diagnostic; vaccine; virucide drug screening;
KW HTLV infection; L93-79C cell culture.
OS Simian T-cell lymphotropic virus (STLVpan-p).
FH Key Location/Qualifiers
FT protein 1..22 /note= "amino acids encoded by T05116"
FT protein 23..171 /note= "amino acids encoded by T05117"
FT protein 40..78 /note= "STLVpan-p rex fragment (see R84655)"
FT W09529240-A1.
PN 02-NOV-1995.
PD 21-APR-1995; U04910.
PR 22-APR-1994; US-231526.
PA (USSH ) US SEC DEPT HEALTH.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Franchini G, Gallo RC, Giri A, Markham P;
DR WPI: 95-382988/49.
DR N-PSDB; T05116, T05117.
PT Isolation and characterisation of primate T-cell lymphotropic virus
PT - used in diagnostic assays and vaccines and to prevent and treat
PT STLvpan-p viral infection in mammals.
PS Disclosure; Page 25; 89pp; English.
CC The sequence represents the STLvpan-p rex protein. Virus
CC non-structural or structural proteins, prepared by recombinant
CC DNA methods, may be used in diagnostic and vaccine applications.
CC Antibodies to the proteins or the virus may be used (i) to treat
CC virus infections, or (ii) in immunoassays to detect STLvpan-p
CC antigens. Tissue culture systems propagating STLvpan-p can be
CC used to screen for anti-STLVpan-p agents. DNA corresponding to
CC this protein may be used as a primer in the isolation of
CC STLvpan-p using the PCR (see primers T05122-24).
CC Sequence 171 AA;
SQ

```

```
Query Match 39.0%; Score 53; DB 16; Length 171;
Best Local Similarity 50.0%; Pred. No. 2.06e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 15 srpbtptfvgldr 30
QY 336 PDQQPIPTETQVADR 351

RESULT 11
ID W15273 standard; Protein; 240 AA.
AC W15273;
DT 04-AUG-1997 (first entry)
DE Salmonella tyrosine phosphatase StpA.
KW StpA: tyrosine phosphatase; phosphorylation; signal transduction;
KW Salmonella secreted protein; Ssp; bacterial-mediated endocytosis.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT protein 1..240
FT /note= "the amino acid sequence does not
FT correspond to the translated sequence
FT of the StpA nucleic acid of Fig 24"
PN WO9718225-AL.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GCHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR N-PSDB; T67033.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 15; Fig 25; 95pp; English.
CC Salmonella tyrosine phosphatase (StpA) (W15273) is a Salmonella
CC secreted protein (Ssp) that has enzymatic activity. It catalyses
CC the release of phosphate groups from tyrosine residues in proteins.
CC StpA can be used to deactivate or activate eukaryotic or
CC prokaryotic signal transduction proteins that are regulated by the
CC phosphorylation and dephosphorylation of tyrosine residues, thereby
CC altering intracellular signal transduction. Thus, Stp can be used
CC as a research tool to study and evaluate phosphorylation-regulated
CC signal transduction pathways.
SQ Sequence 240 AA;

Query Match 39.0%; Score 53; DB 22; Length 240;
Best Local Similarity 50.0%; Pred. No. 2.06e+02;
Matches 9; Conservative 6; Mismatches 1; Indels 2; Gaps 2;

Db 136 pdhqlpstddqleyladr 153
QY 336 PDQQPIPTETQVADR 351

RESULT 12
ID R62863 standard; Protein; 266 AA.
AC R62863;
DT 14-JUL-1995 (first entry)
DE Korean-type Hepatitis C Virus 897 epitope.
KW Korean-type hepatitis C virus; KHCv; non-A, non-B hepatitis; NANBH;
KW viral protein; antigen; epitope; immunodiagnosis.
OS Hepatitis C Virus (Korean-type).
PN WO9425874-A.
PD 10-NOV-1994.
PF 27-APR-1994; KR0039.
PR 28-APR-1993; KR-007231.
PA (LUCK-) LUCKY CO LTD.
PI Cho JM, Choi DS, Choi DY, Kim CH, Kim IS, So HS;
PI Yang JY;
DR WPI; 94-358478/44.
DR N-PSDB; Q73199.
PT Kit for simultaneous diagnosis of hepatitis B and C - comprising
PT one or more hepatitis B and C virus antigenic proteins including
```

```
one or more epitope(s)
Claim 2; Fig 2; 89pp; English.
The KHCv 897 protein is a preferred antigen for use in an
immunodiagnostic kit for simultaneous detection of hepatitis C
and B viruses. The kit comprises antigenic proteins from both
viruses. The HCV protein is pref. One of KHCv CORE 14,
KHCv 897, KHCv NS4E, KHCv NS4E1E2 or KHCv NS5-1.2 proteins; the
HCV protein is pref. HBV CORE or Pre S2 SAg protein.
SQ Sequence 266 AA;

Query Match 39.0%; Score 53; DB 12; Length 266;
Best Local Similarity 44.4%; Pred. No. 2.06e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Db 23 pspavp-gtfqvalhla 39
QY 336 PDQQPIPTETQVADRHP 353

RESULT 13
ID W89253 standard; Protein; 1274 AA.
AC W89253;
DT 10-MAR-1999 (first entry)
DE Human ALP.
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
OS Homo sapiens.
PN WO9849317-A2.
PD 05-NOV-1998.
PF 27-APR-1998; U08439.
PR 23-OCT-1997; US-063595.
PR 28-APR-1997; US-044428.
PR 20-MAY-1997; US-047222.
PR 11-JUN-1997; US-049477.
PR 11-JUN-1997; US-049756.
PR 18-JUN-1997; US-049914.
PA (SUGE-) SUGEN INC.
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,
PI Markby D, Onrust S, Peles E, Plowman GD;
DR WPI; 99-009434/01.
DR N-PSDB; V81748.
PT New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
PS Claim 2; Page 160-164; 193pp; English.
CC The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents human ALP. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
SQ Sequence 1274 AA;

Query Match 39.0%; Score 53; DB 39; Length 1274;
Best Local Similarity 35.0%; Pred. No. 2.06e+02;
Matches 7; Conservative 6; Mismatches 6; Indels 1; Gaps 1;

Db 550 vppqplptpytpagakqp 569
QY 335 LPDQPIPTETQVADRHP 353
```

Search completed: Thu Jul 8 18:48:28 1999
Job time : 19 secs.

```
RESULT 14
ID R30616 standard; Protein; 3010 AA.
AC R30616;
DT 19-MAY-1993 (first entry)
DE Polypeptide coded by Korean HCV full cDNA sequence LBCL.
KW KHCY-LBCL; diagnosis; vaccine.
OS Korean hepatitis C virus.
PN EP-521318-A.
PD 07-JAN-1993.
PF 10-JUN-1992; 109753.
PR 10-JUN-1991; KR-009510.
PR 06-AUG-1991; KR-013601.
PA (LUCK-) LUCKY LTD.
PI Cho JM, Choi DY, Kim CH, Kim ST, Lee YB, Lim KJ, Park YW,
PI So HS, Yang JY.
DR WPI: 93-001883/01.
DR N-PSDB: Q33282.
PT DNA and polypeptide(s) from a new type of hepatitis C virus (KHCY)
PT - for diagnosing and vaccinating against KHCY infections
PS Disclosure: Fig 2: 119pp; English.
CC The polypeptide is that encoded by the full cDNA sequence of Korean
CC hepatitis C virus (KHCY) cDNA, KHCY-LBCL. It or its fragments may be
CC used in a specific and accurate method for detecting KHCY antibodies
CC in the serum of hepatitis C patients. Antibodies directed against these
CC polypeptides are useful for the purification of KHCY antigens and for
CC the development of an improved diagnostic to detect KHCY antigens in a
CC sample. The polypeptides may also be used in a vaccine for treatment
CC and prevention of KHCY infection at a dosage of 5-200 ug/peptide.
SQ Sequence 3010 AA;

Query Match 39.0%; Score 53; DB 6; Length 3010;
Best Local Similarity 44.4%; Pred. No. 2.06e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Db 1214 pspavp-qtfqvalha 1230
Qy 336 PDQPIPTETQVADRRP 353

RESULT 15
ID W81969 standard; Protein; 318 AA.
AC W81969;
DT 05-FEB-1999 (first entry)
DE Mouse EI24 protein #2.
KW EI24; etoposide-induced apoptosis; murine; degenerative disorder;
KW cell proliferation; cell death; p53.
OS Mus sp.
PN US5843659-A.
PD 01-DEC-1998.
PF 21-MAR-1996; 619362.
PR 21-MAR-1996; US-619362.
PA (APOPT-) APOPTOSIS TECHNOLOGY INC.
PI Guild BC, Lehar SM;
DR WPI: 99-044568/04.
PT Cloned etoposide-induced apoptosis gene EI24 - useful for diagnosing
PT degenerative disorders characterised by inappropriate cell
PT proliferation or death
PT Example; Fig 3; 35pp; English.
CC This sequence represents a murine etoposide-induced apoptosis gene, EI24.
CC The EI24 gene product can be used in the diagnosis of degenerative
CC disorders characterised by inappropriate cell proliferation or death.
CC Induction of this gene by etoposide requires expression of wild-type
CC p53.
SQ Sequence 318 AA;

Query Match 38.2%; Score 52; DB 38; Length 318;
Best Local Similarity 40.0%; Pred. No. 2.55e+02;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 166 qdudkafevsgrkp 180
Qy 339 OPIPTETQVADRRP 353
```

WQSRH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:47:36 1999; MasPar time 3 47 Seconds
Tabular output not generated. 219.085 Million cell updates/sec

Title: >US-09-041-236-2
Description: (335-353) from US09041236.ppep (20 of 45)
Perfect Score: 136
Sequence: 1 LPDQPIPTTFQVADRRHP 19

Scoring table: PAM 150
Gap 11
Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1.pir1 2.pir2 3.pir3 4.pir4

Statistics: Mean 27.437; Variance 41.176; scale 0.666

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	95	69.9	653	2	T03102 semaphorin homolog A3	9.17e+07
2	72	52.9	299	2	B69988 hypothetical protein	3.88e-02
3	63	46.3	665	2	S70706 probable protein kina	1.68e+00
4	63	46.3	815	2	B56708 extracellular signal-	1.68e+00
5	61	44.9	337	2	F70323 transcription termina	3.71e+00
6	60	44.1	213	1	JMB0 amlogenin I precurs	5.49e+00
7	60	44.1	250	2	S62838 hypothetical protein	5.49e+00
8	60	44.1	1458	2	S36014 dynein heavy chain, c	5.49e+00
9	59	43.4	529	2	S68120 lactase (EC 1.10.3.2)	8.09e+00
10	59	43.4	549	2	S39585 peptide transport per	8.09e+00
11	59	43.4	723	1	RMQTN RNA-directed RNA poly	8.09e+00
12	58	42.6	148	2	S07196 me13 protein - fissio	1.19e+01
13	58	42.6	379	2	T01034 sulfotransferase hom	1.19e+01
14	58	42.6	387	2	S52274 transcription termina	1.19e+01
15	58	42.6	541	2	S42833 F40f12.4 protein - Ca	1.19e+01
16	58	42.6	610	2	S76234 hypothetical protein	1.19e+01
17	58	42.6	619	2	T02922 nikB protein - Escher	1.19e+01
18	57	41.9	173	2	S41755 cyclin E type II - fr	1.73e+01
19	57	41.9	408	2	S42493 opaque 2 protein - Jo	1.73e+01
20	56	41.2	184	2	S26046 sex-determining prote	2.52e+01
21	56	41.2	296	2	JW0103 azarene carbazole ca	2.52e+01
22	56	41.2	457	2	S44269 platelet-derived grow	2.52e+01
23	56	41.2	477	1	B46396 transcription factor	2.52e+01

24	56	41.2	543	2	B40033 protein kinase (EC 2.	2.52e+01
25	56	41.2	907	2	A24938 hypothetical T2 prote	2.52e+01
26	55	40.4	195	2	C70928 probable transcriptio	3.64e+01
27	55	40.4	458	2	I39477 glutathione reductase	3.64e+01
28	55	40.4	561	2	S62552 hypothetical protein	3.64e+01
29	55	40.4	761	2	S66719 hypothetical protein	3.64e+01
30	55	40.4	972	1	GNXSIV genome polyprotein -	3.64e+01
31	55	40.4	2406	2	A54148 odz protein - fruit f	3.64e+01
32	55	40.4	2515	2	S47008 tenascin-like protein	3.64e+01
33	54	39.7	192	2	F41816 amelogenin (Y chromo	5.23e+01
34	54	39.7	201	1	Q0BP22 orf-201 protein - pha	5.23e+01
35	54	39.7	257	2	D70552 hypothetical protein	5.23e+01
36	54	39.7	357	2	S18236 omega secalin precurs	5.23e+01
37	54	39.7	357	2	S18235 omega secalin precurs	5.23e+01
38	54	39.7	398	2	A40966 porphobilinogen synh	5.23e+01
39	54	39.7	458	2	S75114 hypothetical protein	5.23e+01
40	54	39.7	601	1	B56564 transcription factor	5.23e+01
41	54	39.7	669	2	I38029 matrix metalloprotein	5.23e+01
42	54	39.7	834	2	S54563 hypothetical protein	5.23e+01
43	54	39.7	899	2	B38529 nikB protein - Escher	5.23e+01
44	54	39.7	925	2	S27920 nuclear antigen EBNA-	5.23e+01
45	54	39.7	2616	2	A57096 nudel protein precurs	5.23e+01

ALIGNMENTS

RESULT 1
ENTRY T03102 #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS T03102
REFERENCE Z14840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
#status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-653 ##label ENS
##cross-references EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY #length 653 #molecular-weight 73645 #checksum 5501
Query Match 69.9%; Score 95; DB 2; Length 653;
Best Local Similarity 63.2%; Pred. No. 9.17e+07;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 389 MKNHQVPTETFOVADRRYP 407
Qy 335 LPDQPIPTTFQVADRRHP 353

RESULT 2
ENTRY B69988 #type complete
TITLE hypothetical protein ytap - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
ACCESSIONS B69988
REFERENCE A69880
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;

Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoett, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppe, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,

C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kashahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueil, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Ogestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
#cross-references MUID:98044033
#accession B69988
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-299 ##label KUN
##cross-references GB:299119; GB:AL009126; NID:g2635411; PID:el185988;
PID:g2635509
##experimental_source strain 168

GENETICS
#gene ytaP
SUMMARY
#length 299 #molecular-weight 33776 #checksum 6820
Query Match 52.9%; Score 72; DB 2; Length 299;
Best Local Similarity 47.1%; Pred. No. 3.88e-02;
Matches 8; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 16 LPDRRPISVETLRIER 32
|||:||||:||||:|
QY 335 LPDQPIPTTFQVADR 351

RESULT 3
ENTRY S70706 #type complete
TITLE probable protein kinase TB3 (EC 2.7.1.-) - Colletotrichum trifolii
ALTERNATE_NAMES protein kinase COT1 homolog
ORGANISM #formal_name Colletotrichum trifolii
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Jul-1998
ACCESSIONS S70706
REFERENCE S70706
#authors Buhr, T.L.; Oved, S.; Truesdell, G.M.; Huang, C.; Yarden, O.; Dickman, M.B.
#journal Mol. Gen. Genet. (1996) 251:565-572
#title A kinase-encoding gene from *Colletotrichum trifolii* complements a colonial growth mutant of *Neurospora crassa*.
#cross-references MUID:96323031
#accession S70706
##molecule_type DNA
##residues 1-665 ##label BUH
##cross-references EMBL:U14989; NID:g641978; PID:g641979

GENETICS
#gene TB3
#introns 236/3; 404/2; 562/2
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology

C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kashahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueil, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Ogestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
#cross-references MUID:98044033
#accession B69988
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-299 ##label KUN
##cross-references GB:299119; GB:AL009126; NID:g2635411; PID:el185988;
PID:g2635509
##experimental_source strain 168

GENETICS
#gene ytaP
SUMMARY
#length 299 #molecular-weight 33776 #checksum 6820
Query Match 52.9%; Score 72; DB 2; Length 299;
Best Local Similarity 47.1%; Pred. No. 3.88e-02;
Matches 8; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 16 LPDRRPISVETLRIER 32
|||:||||:||||:|
QY 335 LPDQPIPTTFQVADR 351

RESULT 3
ENTRY S70706 #type complete
TITLE probable protein kinase TB3 (EC 2.7.1.-) - Colletotrichum trifolii
ALTERNATE_NAMES protein kinase COT1 homolog
ORGANISM #formal_name Colletotrichum trifolii
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Jul-1998
ACCESSIONS S70706
REFERENCE S70706
#authors Buhr, T.L.; Oved, S.; Truesdell, G.M.; Huang, C.; Yarden, O.; Dickman, M.B.
#journal Mol. Gen. Genet. (1996) 251:565-572
#title A kinase-encoding gene from *Colletotrichum trifolii* complements a colonial growth mutant of *Neurospora crassa*.
#cross-references MUID:96323031
#accession S70706
##molecule_type DNA
##residues 1-665 ##label BUH
##cross-references EMBL:U14989; NID:g641978; PID:g641979

GENETICS
#gene TB3
#introns 236/3; 404/2; 562/2
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology

phosphotransferase; protein kinase
FEATURE
44-84 #region glutamine-rich\
280-585 #domain protein kinase homology #label KIN
SUMMARY #length 665 #molecular-weight 76015 #checksum 6024
Query Match 46.3%; Score 63; DB 2; Length 665;
Best Local Similarity 36.8%; Pred. No. 1.68e-00;
Matches 7; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 183 LPNQPASVDPPAPAPER 201
||:||||:|:|:
QY 335 LPDQPIPTTF-QVADRH 352

RESULT 4
ENTRY B56708 #type complete
TITLE extracellular signal-regulated kinase 5 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 19-Dec-1997
ACCESSIONS B56708
REFERENCE A56708
#authors Zhou, G.; Bao, Z.Q.; Dixon, J.E.
#journal J. Biol. Chem. (1995) 270:12665-12669
#title Components of a new human protein kinase signal transduction pathway.
#cross-references MUID:95279403
#accession B56708
##status preliminary
##molecule_type mRNA
##residues 1-815 ##label ZHO
##cross-references GB:U25278; NID:g837260; PID:g837261
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology
ATP

KEYWORDS
FEATURE
52-315 #domain protein kinase homology #label KIN\
60-68 #region protein kinase ATP-binding motif
SUMMARY #length 815 #molecular-weight 88636 #checksum 7383
Query Match 46.3%; Score 63; DB 2; Length 815;
Best Local Similarity 50.0%; Pred. No. 1.68e-00;
Matches 10; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 300 LPPRQVPWETVPGADROA 319
||:||||:|:|:
QY 335 LPDQPIPTTF-FQVADRH 353

RESULT 5
ENTRY F70323 #type complete
TITLE transcription termination Nua - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 21-Aug-1998
ACCESSIONS F70323
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.
#cross-references MUID:98196666
#accession F70323
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-337 ##label AQF
##cross-references GB:AE000681; NID:g2982963; PID:g2982967; GB:AE000657
##experimental_source strain Vfs

REFERENCE	PC2005	'Q', 170-171, 176, 'F', 178-182, 187-194, 'N', 196-198, 'Q', 200-201 ##label TAK
#authors	Fincham, A.G.; Moradian-Oldak, J.	
#journal	Biochem. Biophys. Res. Commun.	(1993) 197:248-255
#title	Amelogenin post-translational modifications: carboxy-terminal processing and the phosphorylation of bovine and porcine "TRAP" and "LRAP" amelogenins.	
#cross-references	MUID:94071951	
#accession	PC2005	
##molecule_type	protein	
##residues	17-61	##label FI2
#accession	PC2007	
##molecule_type	protein	
##residues	17-40; 48-49, 188-202	##label FIN
COMMENT	Amelogenin is the predominant protein in developing dental enamel.	
GENETICS		
#map_position	X	
CLASSIFICATION	superfamily amelogenin	
KEYWORDS	alternative splicing; enamel; phosphoprotein; tandem repeat; tooth	
FEATURE		
1-16	#domain signal sequence #status predicted #label SIG\	
17-213	#product amelogenin #status predicted #label MAT\	
17-61	#product tyrosine-rich amelogenin peptide, TRAP-45	
	#status experimental #label MA2\	
17-59	#product tyrosine-rich amelogenin peptide, TRAP-43	
	#status experimental #label MA3\	
17-49, 188-213	#product leucine-rich amelogenin peptide, LRAP-59	
	#status experimental #label MA4\	
17-49, 188-202	#product leucine-rich amelogenin peptide, LRAP-48	
	#status experimental #label MA5\	
17-49, 188-196	#product leucine-rich amelogenin peptide, LRAP-43	
	#status experimental #label MA6\	
136-165	#region 3-residue repeats (Q-P-X)\	
32	#binding_site phosphate (Ser) (covalent) #status predicted	
SUMMARY	#length 213	#molecular-weight 24119 #checksum 90
Query Match	44.1%	Score 60; DB 1; Length 213;
Best Local Similarity	36.8%	Pred. No. 5.49e+00;
Matches	7; Conservative	6; Mismatches 6; Indels 0; Gaps 0;
Db	122 LPAQQFPQSIQPPHPQ 140	
	: : :	
QY	335 LPDQPIPTTFQVADHP 353	
RESULT	7	
ENTRY	S62838	#type complete
TITLE	hypothehtical protein K05_orf250 - Mycoplasma pneumoniae (ATCC 29342) (SGC3)	
ORGANISM	#formal_name Mycoplasma pneumoniae	
#variety	ATCC 29342	
DATE	24-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 18-Sep-1998	
ACCESSIONS	S62838: S73481	
REFERENCE	S62797	
#authors	Hilbert, H.; Himmelreich, R.; Plagens, H.; Herrmann, R.	
#journal	Nucleic Acids Res.	(1996) 24:628-639
#title	Sequence analysis of 56 kb from the genome of the bacterium Mycoplasma pneumoniae comprising the dnaA region, the atp operon and a cluster of ribosomal protein genes.	
#cross-references	MUID:96177562	
#accession	S62838	
##status	preliminary; nucleic acid sequence not shown; translation not shown	
##molecule_type	DNA	
##residues	1-250	##label HIL
##cross-references	EMBL:U34816; NID:g1209514; PTD:g1209519	
##note	the nucleotide sequence was submitted to the EMBL Data Library, August 1995	
REFERENCE	S73327	

```
#authors Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li,
#journal B.C.; Herrmann, R.
#title Nucleic Acids Res. (1996) 24:4420-4449
#cross-references Complete sequence analysis of the genome of the bacterium
#accession Mycoplasma pneumoniae.
#status MUID:97105885
#molecule_type preliminary: nucleic acid sequence not shown:
#residues translation: not shown
#cross-references EMBL:AE000017; GB:U00089; NID:gl673812; PID:gl673816
#note the nucleotide sequence was submitted to the EMBL Data
Library, November 1996
GENETICS
#genetic_code SGC3
SUMMARY #length 250 #molecular-weight 28662 #checksum 6202
Query Match 44.1%; Score 60; DB 2; Length 250;
Best Local Similarity 41.2%; Pred. No. 5.49e+00;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
Db 139 NQPVATSNIOVEATQP 155
:||||:| :|| :|
Qy 337 DQOQPIPTETFOVADRRHP 353
:||||:| :|| :|
RESULT 8
ENTRY S36014 #type fragment
TITLE dynein heavy chain, cytosolic - sea urchin (Tripneustes
ORGANISM gratilla) (fragment)
#formal_name Tripneustes gratilla
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
08-Sep-1997
ACCESSIONS S36014
REFERENCE Gibbons, I.
#authors submitted to the EMBL Data Library, March 1993
#accession S36014
#molecule_type mRNA
#residues 1-1458 #label GIB
#cross-references EMBL:Z21941; NID:9287739; PID:g758347
CLASSIFICATION #superfamily dynein heavy chain, cytosolic
KEYWORDS cytosol
SUMMARY #length 1458 #checksum 8163
Query Match 44.1%; Score 60; DB 2; Length 1458;
Best Local Similarity 46.7%; Pred. No. 5.49e+00;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Db 951 LPPQQTVPIDFEVS 965
:||||:| :|| :|
Qy 335 LPDQQPIPTETFOVA 349
:||||:| :|| :|
RESULT 9
ENTRY S68120 #type complete
TITLE lactase (EC 1.10.3.2) 4 precursor - Rhizoctonia solani
ORGANISM #formal_name Rhizoctonia solani
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
13-Sep-1998
ACCESSIONS S68120; S77976; S72564
REFERENCE Wahleithner, J.A.; Xu, F.; Brown, K.M.; Brown, S.H.;
#authors Golightly, E.J.; Halkier, T.; Kauppinen, S.; Pederson, A.;
Schneider, P.
#journal Curr. Genet. (1996) 29:395-403
#title The identification and characterization of four lactases from
the plant pathogenic fungus Rhizoctonia solani.
#cross-references MUID:96171523
#accession S68120
#molecule_type DNA
#residues 1-529 #label WAHL
```

```
#accession S77976
#molecule_type protein
#residues 20-53;132-180;203-217;378-396;440-465;489-513;520-529
#label WAH2
REFERENCE S72564
#authors Wahleithner, J.A.
#submission submitted to the EMBL Data Library, September 1995
#accession S72564
#molecule_type DNA
#residues 1-417,'GAQHHR',424-469,'A',471-483,'CT',484,'R',486,
'GSH',490,'S',492,'PRPP',497,'RRARVSSRS',499,'PMP',
504,'TS',507,'APSTRFLPLI',518,'S',520,'W' ##label WAH
#cross-references EMBL:254277; NID:gl1150567; PID:e206168; PID:gl1150568
GENETICS
#gene lcc4
#introns 55/3; 70/3; 82/3; 87/3; 101/1; 126/2; 143/1; 147/1; 170/2;
193/2; 250/3; 276/3; 303/1; 340/3; 379/3; 417/3; 470/1
#feature copper; oxidoreductase
KEYWORDS #domain signal sequence #status predicted #label SIG\
FEATURE 1-19 #product lactase 4 #status experimental #label MAT
20-529 #length 529 #molecular-weight 57301 #checksum 1923
SUMMARY
Query Match 43.4%; Score 59; DB 2; Length 529;
Best Local Similarity 45.5%; Pred. No. 8.09e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 257 HQPLPVDSEFI 267
:||||:| :|| :|
Qy 338 QOPIPTETFOV 348
:||||:| :|| :|
RESULT 10
ENTRY S39585 #type complete
TITLE peptide transport periplasmic protein sapA precursor -
ORGANISM Salmonella typhimurium
#formal_name Salmonella typhimurium
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
05-Dec-1998
ACCESSIONS S39585
REFERENCE Parra-Lopez, C.; Baer, M.T.; Groisman, E.A.
#authors EMBO J. (1993) 12:4053-4062
#journal Molecular genetic analysis of a locus required for resistance
#title to antimicrobial peptides in Salmonella typhimurium.
#cross-references MUID:94038887
#accession S39585
#status preliminary
#molecule_type DNA
#residues 1-549 #label PAR
#cross-references EMBL:X74212; NID:g414207; PID:g414208
#experimental_source strain 14028s
GENETICS
#gene sapA
CLASSIFICATION #superfamily dipeptide transport protein
KEYWORDS periplasmic space; transport system
FEATURE 1-21 #domain signal sequence #status predicted #label SIG\
22-549 #product peptide transport periplasmic protein sapA
#status predicted #label MAT
SUMMARY #length 549 #molecular-weight 61622 #checksum 895
Query Match 43.4%; Score 59; DB 2; Length 549;
Best Local Similarity 43.8%; Pred. No. 8.09e+00;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Db 209 LLDROPVGTGPFQLSE 224
:||||:| :|| :|
Qy 335 LPDQQPIPTETFOVAD 350
:||||:| :|| :|
RESULT 11
ENTRY RRWQTN #type complete
```

```

TITLE      RNA-directed RNA polymerase (EC 2.7.7.48) - tobacco necrosis
            virus (strain A)
ALTERNATE_NAMES  RNA nucleotidyltransferase (RNA-directed): RNA replicase
CONTAINS        23k protein
ORGANISM        #formal_name tobacco necrosis virus, TNV
                #note host Phaseolus vulgaris (kidney bean)
                DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
                    05-Sep-1997
ACCESSIONS     A35523
REFERENCE      A35523
#authors      Meulwaeter, F.; Seurinck, J.; van Emmelo, J.
#journal      Virology (1990) 177:699-709
#title        Genome structure of tobacco necrosis virus strain A.
#cross-references  MUID:90320143
#accession    A35523
#molecule_type genomic RNA
#residues     1-723 #label MEU
#cross-references EMBL:M33002; NID:g1172549; PID:g310915
#note         readthrough of the terminator UAG occurs between codons
                AAA for 202-Lys and GGG for 203-Gly
CLASSIFICATION #superfamily carnation mottle virus RNA-directed RNA
                polymerase; barley yellow dwarf virus RNA-directed RNA
                polymerase homology
KEYWORDS       nucleotidyltransferase; RNA biosynthesis; RNA replication
FEATURE       1-202 #product 23k protein #status predicted #label KPP\
427-593        #domain barley yellow dwarf virus RNA-directed RNA
                polymerase homology #label BYD
SUMMARY        #length 723 #molecular-weight 82167 #checksum 2224
                Query Match 43.4%; Score 59; DB 1; Length 723;
                Best Local Similarity 58.3%; Pred. No. 8.09e+00;
                Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 705 IPTDTYQVSGEH 716
QY 341 IPTETFOVADR 352

RESULT 12
ENTRY   S07196 #type complete
TITLE   me13 protein - fission yeast (Schizosaccharomyces pombe)
ORGANISM #formal_name Schizosaccharomyces pombe
DATE     30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
                    09-Sep-1997
ACCESSIONS S07196
REFERENCE  S07196
#authors   McLeod, M.; Stein, M.; Beach, D.
#journal   EMBO J. (1987) 6:729-736
#title     The product of the me13+ gene, expressed under control of the
            mating-type locus, induces meiosis and sporulation in
            fission yeast.
#cross-references MUID:87218534
#accession S07196
#molecule_type DNA
#residues 1-148 #label MCL
#cross-references EMBL:X05142; NID:g4992; PID:g4993
#gene      me13
SUMMARY    #length 148 #molecular-weight 16473 #checksum 6745
                Query Match 42.6%; Score 58; DB 2; Length 148;
                Best Local Similarity 58.3%; Pred. No. 1.19e+01;
                Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 131 PRPNETFDITDR 142
QY 340 PIPTETFOVADR 351

RESULT 13
ENTRY   T01034 #type complete
TITLE   sulfurtransferase homolog YUP8H12R.17 - Arabidopsis thaliana

```

```

ORGANISM    #formal_name Arabidopsis thaliana #common_name mouse-ear
            cress
DATE        05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change
            05-Feb-1999
ACCESSIONS  T01034
REFERENCE    214227
#authors     Theologis, A.; Vysotskaia, V.S.; Osborne, B.I.; Schwartz,
            J.R.; Federspiel, N.A.; Kwan, A.; Toriumi, M.; Yu, G.; Oji,
            O.; Araujo, R.; Chung, E.; Dewar, K.; Dietrich, F.; Ecker,
            J.R.; Marziani, A.; Oefner, P.; Davis, R.W.
            #submission submitted to the EMBL Data Library, May 1998
            #description Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.
            #accession T01034
            #status preliminary; translated from GB/EMBL/DDBJ
            #molecule_type DNA
            #residues 1-379 #label THE
            #cross-references EMBL:AC002986; NID:g2494106; PID:g3152581
GENETICS     #map_position 1
            #introns 34/3; 96/3; 116/3; 137/3; 179/2; 233/3; 253/3; 273/2; 300/3;
            322/1; 344/3
            #note YUP8H12R.17
            SUMMARY #length 379 #molecular-weight 41893 #checksum 5548
                Query Match 42.6%; Score 58; DB 2; Length 379;
                Best Local Similarity 46.7%; Pred. No. 1.19e+01;
                Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
Db 104 MPDEORNPTOEYQVA 118
QY 335 LPDQOPIPTETFOVA 349

RESULT 14
ENTRY   S52274 #type complete
TITLE   transcription termination-antitermination factor nusA -
            Thermus aquaticus
ORGANISM #formal_name Thermus aquaticus
DATE     08-May-1995 #sequence_revision 21-Jul-1995 #text_change
            30-Jun-1998
ACCESSIONS S52274
REFERENCE  S52273
            Vornlocher, H.; Sprinzl, M.
            #authors submitted to the EMBL Data Library, January 1995
            #description Molecular cloning of the Thermus thermophilus nusA/infB
            operon.
            #accession S52274
            #status preliminary
            #molecule_type DNA
            #residues 1-387 #label VOR
            #cross-references EMBL:248001; NID:g642362; PID:g642364
            #note the source is designated as Thermus thermophilus
            #superfamily Bacillus transcription termination factor nusA;
            #classification transcription termination factor nusA homology
FEATURE     74-349
            #domain transcription termination factor nusA homology
            #label TTN
            SUMMARY #length 387 #molecular-weight 43939 #checksum 6179
                Query Match 42.6%; Score 58; DB 2; Length 387;
                Best Local Similarity 41.2%; Pred. No. 1.19e+01;
                Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
Db 172 LPRSEOIPTKEVHPGOR 188
QY 335 LPDQOPIPTETFOVADR 351

RESULT 15
ENTRY   S42833 #type complete
TITLE   F40F12.4 protein - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE     20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change

```

09-Sep-1997
ACCESSIONS S42833
REFERENCE S42830
#authors Kershaw, J.
#submission Submitted to the EMBL Data Library, February 1994
#accession S42833
##status preliminary
##molecule_type DNA
##residues 1-541 #label KER
##cross-references EMBL:230215; NID:g456511; PID:g456515
GENETICS
#introns 21/1: 45/1: 91/1: 153/1: 190/2: 258/3: 344/3: 425/1: 459/3:
507/3
SUMMARY #length 541 #molecular-weight 59963 #checksum 3718
Query Match 42.6%; Score 58; DB 2: Length 541;
Best Local Similarity 35.3%; Pred. No. 1.19e+01;
Matches 6; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
Db 205 IPEQGPVPFIYIIADH 221
:|:|:|:|:|:|:
Qy 335 LPDQOPIPTETQVADR 351

Search completed: Thu Jul 8 18:47:49 1999
Job time : 13 secs.

WQRLH (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:46:32 1999; Maspar time 2.52 Seconds
Tabular output not generated. 213.225 Million cell updates/sec

Title: >US-09-041-236-2
Description: (335-353) from US09041236.pep (20 of 45)
Perfect Score: 136
Sequence: 1 LPDQDPIPTETQVADRH 19

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 28.335; Variance 37.196; scale 0.762

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	63	46.3	815	1	ERK5_HUMAN	5.39e+01
2	60	44.1	213	1	AMEX_BOVIN	2.01e+00
3	59	43.4	531	1	LAC4_THACU	3.08e+00
4	59	43.4	549	1	SAPA_SALTY	3.08e+00
5	59	43.4	723	1	RPO_TNVA	3.08e+00
6	58	42.6	148	1	MEI3_SCHPO	3.08e+00
7	58	42.6	387	1	NUSA_THETH	4.70e+00
8	58	42.6	469	1	PAF_ARATH	4.70e+00
9	58	42.6	971	1	Y029_HUMAN	4.70e+00
10	57	41.9	289	1	HOG3_HORVU	7.14e+00
11	57	41.9	548	1	YD03_SCHPO	7.14e+00
12	56	41.2	184	1	TRSF_DROSI	1.08e+01
13	56	41.2	477	1	ETV1_MOUSE	1.08e+01
14	56	41.2	720	1	ERK3_RAT	1.08e+01
15	56	41.2	815	1	EBN3_EBV	1.08e+01
16	55	40.4	195	1	YX44_MYCTU	1.62e+01
17	55	40.4	214	1	YDL8_SCHPO	1.62e+01
18	55	40.4	392	1	YTM_RHISN	1.62e+01
19	55	40.4	422	1	SPM1_SCHPO	1.62e+01
20	55	40.4	458	1	GHR_ANASP	1.62e+01
21	55	40.4	561	1	YAI7_SCHPO	1.62e+01
22	55	40.4	626	1	Z189_HUMAN	1.62e+01
23	55	40.4	741	1	RED1_HUMAN	1.62e+01

24	55	40.4	972	1	POLS_IPNVJ	STRUCTURAL POLYPROTEIN	1.62e+01
25	55	40.4	1790	1	USOL_YEAST	INTRACELLULAR PROTEIN	1.62e+01
26	54	39.7	155	1	TL20_NPPOP	TELOKIN-LIKE PROTEIN-2	2.41e+01
27	54	39.7	189	1	AMEL_PIG	AMELOGENIN PRECURSOR (2.41e+01
28	54	39.7	192	1	AMEY_HUMAN	AMELOGENIN, Y ISOFORM	2.41e+01
29	54	39.7	201	1	YG15_BPP22	HYPOTHETICAL 23.8 KD P	2.41e+01
30	54	39.7	398	1	HEM2_BEA	DELTA-AMINOLEVULINIC A	2.41e+01
31	54	39.7	412	1	HEM2_SOYBN	DELTA-AMINOLEVULINIC A	2.41e+01
32	54	39.7	420	1	Y4ME_RHISN	HYPOTHETICAL 45.7 KD P	2.41e+01
33	54	39.7	601	1	PDML_DROME	NUBBIN PROTEIN (TWIN)	2.41e+01
34	54	39.7	604	1	PGH2_RABIT	PROSTAGLANDIN G/H SYNT	2.41e+01
35	54	39.7	669	1	COGU_HUMAN	MATRIX METALLOPROTEIN	2.41e+01
36	54	39.7	686	1	IF2M_SCHPO	TRANSLATION INITIATION	2.41e+01
37	54	39.7	834	1	YM22_YEAST	HYPOTHETICAL 94.3 KD T	2.41e+01
38	54	39.7	2616	1	NDL_DROME	SERINE PROTEASE NUDEL	2.41e+01
39	53	39.0	99	1	RS10_HALMA	30S RIBOSOMAL PROTEIN	3.57e+01
40	53	39.0	185	1	ADML_RAT	ADRENOMEDULLIN PRECURS	3.57e+01
41	53	39.0	448	1	YKV3_YEAST	HYPOTHETICAL 51.6 KD P	3.57e+01
42	53	39.0	477	1	ETV1_HUMAN	ETS TRANSLOCATION VARI	3.57e+01
43	53	39.0	540	1	YMA5_CAEEL	HYPOTHETICAL ZINC FING	3.57e+01
44	53	39.0	762	1	MUS2_AQUAE	MUTS2 PROTEIN.	3.57e+01
45	53	39.0	1025	1	YESO_ECOLI	HYPOTHETICAL 111.0 KD	3.57e+01

ALIGNMENTS

RESULT 1	ID	ERK5_HUMAN	STANDARD;	PRT;	815 AA.
AC	Q13164;	Q16634;			
DT	01-NOV-1997	(REL. 35, CREATED)			
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)			
DE	EXTRACELLULAR SIGNAL-REGULATED KINASE 5 (EC 2.7.1.-) (ERK5) (ERK4)				
DE	(BMK1 KINASE).				
GN	ERK5 OR ERK4.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;				
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=FETAL BRAIN;				
RX	MEDLINE; 95279403.				
RA	ZHOU G.; BAO Z.G.; DIXON J.E.;				
RT	"Components of a new human protein kinase signal transduction pathway."				
RL	J. BIOL. CHEM. 270:12665-12669(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=PLACENTA;				
RX	MEDLINE; 95374539.				
RA	LEE J.-D.; ULEVITCH R.J.; HAN J.;				
RT	"Primary structure of BMK1: a new mammalian map kinase.";				
RL	BIOCHEM. BIOPHYS. RES. COMMUN. 213:715-724(1995)				
CC	-!- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.				
CC	-!- ENZYME REGULATION: ACTIVATED BY TYROSINE AND THREONINE PHOSPHORYLATION (BY SIMILARITY).				
CC	-!- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE IN LIVER.				
CC	-!- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.				
CC	-!- PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY ROLE, IS ABSENT.				
CC	-!- SIMILARITY: BELONGS TO THE CDC2/CDC28 SUBFAMILY OF SER/THR PROTEIN KINASES. STRONGEST SIMILARITY WITH OTHER MAP KINASES.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial				

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
-----
CC  ENBL; U25278: G837261; -
CC  ENBL; U29725: G973307; -
CC  ENBL; U29725: G973309; -
CC  ENBL; U29727: G973311; -
CC  DR PROSITE; PS00107: PROTEIN_KINASE_ATP; 1.
CC  DR PROSITE; PS00108: PROTEIN_KINASE_ST; 1.
CC  DR PROSITE; PS50011: PROTEIN_KINASE_DOM; 1.
CC  PFAM; PF00089; pkinase; 1.
CC  HSP; P24941; IAOI.
CC  TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING; CELL CYCLE;
KW PHOSPHORYLATION.
FT FT DOMAIN 54 346 PROTEIN KINASE.
FT NP_BIND 60 68 ATP (BY SIMILARITY).
FT BINDING 83 83 ATP (BY SIMILARITY).
FT ACT_SITE 181 181 BY SIMILARITY.
FT MOD_RES 218 218 PHOSPHORYLATION (ACTIVATES THE KINASE)
FT MOD_RES 220 220 PHOSPHORYLATION (ACTIVATES THE KINASE)
FT DOMAIN 337 340 POLY-ALA.
FT DOMAIN 433 464 PRO-RICH 1.
FT DOMAIN 520 523 POLY-ARG.
FT DOMAIN 577 699 PRO-RICH 2.
FT CONFLICT 19 32 AREGTRPRHCLCS -> GPVKVEPAHTAASVA (IN REF. 2).
FT CONFLICT 609 609 L -> V (IN REF. 2).
FT SEQUENCE 815 AA; 88636 MW; C54AB952 CRC32;

Query Match 46.3%; Score 63; DB 1; Length 815;
Best Local Similarity 50.0%; Pred. No. 5.39e-01;
Matches 10; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 300 LPPRPVPMETVPCADRQA 319
QY 335 LPDQPIPTET-FQVADHRP 353

RESULT 2 STANDARD; PRT; 213 AA.
AC P02817;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE AMELOGENIN, CLASS 1 PRECURSOR.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91113686.
RA GIBSON C., GOLUB E., HEROLD R., RISSER M., DING W., SHIMOKAWA H.,
RA YOUNG M., TERMINE J., ROSENBLUM J.;
RT "Structure and expression of the bovine amelogenin gene.";
RL BIOCHEMISTRY 30:1075-1079(1991).
RN [2]
RP PRELIMINARY SEQUENCE OF 17-201.
RX MEDLINE; 84231410.
RA TAKAGI T., SUZUKI M., BABA T., MINEGISHI K., SASAKI S.;
RT "Complete amino acid sequence of amelogenin in developing bovine enamel.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 121:592-597(1984).
RN [3]
RP ALTERNATIVE SPLICING (LRAP).
RX MEDLINE; 91144612.
RA GIBSON C.W., GOLUB E., DING W., SHIMOKAWA H., YOUNG M., TERMINE J.,
RA ROSENBLUM J.;
RT "Identification of the leucine-rich amelogenin peptide (LRAP) as the
RT translation product of an alternatively spliced transcript.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 174:1306-1312(1991).
```

```
-----
RN SEQUENCE OF 17-49 AND 188-213.
RP MEDLINE; 94071951.
RX FINCHAM A.G., MORADIAN-OLDAK J.;
RT "Amelogenin post-translational modifications: carboxy-terminal
RT processing and the phosphorylation of bovine and porcine 'TRAP' and
RT 'LRAP' amelogenins.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 197:248-255(1993).
RN [5]
RP STRUCTURE BY NMR.
RX MEDLINE; 90091473.
RA RENUGOPALAKRISHNAN V., PRABHAKARAN M., HUANG S.G., BALASUBRAMANIAM A.,
RA STRAWICH E., GLIMCHER M.J.;
RT "Secondary structure and limited three-dimensional structure of
RT bovine amelogenin.";
RL CONNECT. TISSUE RES. 22:131-138(1989).
CC -!- FUNCTION: PLAYS A ROLE IN THE BIOMINERALIZATION OF TOOTH. SEEMS
CC TO REGULATE THE FORMATION OF CRYSTALLITES DURING THE SECRETORY
CC STAGE OF TOOTH ENAMEL DEVELOPMENT. THOUGHT TO PLAY A MAJOR ROLE IN
CC THE STRUCTURAL ORGANIZATION AND MINERALIZATION OF DEVELOPING
CC ENAMEL.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -!- AMELOGENIN IS THE PREDOMINANT PROTEIN IN DEVELOPING DENTAL ENAMEL.
CC -!- CLASS 1 AMELOGENINS LOCATED ON X CHROMOSOME.
CC -!- ALTERNATIVE PRODUCTS: VARIOUS FORMS ARE PRODUCED BY ALTERNATIVE
CC SPLICING. ONE OF THESE FORMS, LRAP, LACKS THE CENTRAL SECTION OF
CC AMELOGENIN.
CC -!- SIMILARITY: BELONGS TO THE AMALOGENIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

Query Match 44.1%; Score 60; DB 1; Length 213;
Best Local Similarity 36.8%; Pred. No. 2.01e+00;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 122 LPAQPPFPQSIQPPHP 140
QY 335 LPDQPIPTET-FQVADHRP 353

RESULT 3
ID LAC4_THACU STANDARD; PRT; 531 AA.

AC Q02081;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE LACCASE 4 PRECURSOR (EC 1.10.3.2) (BENZEDIOL-OXYGEN OXIDOREDUCTASE)
DE (URISHIOL OXIDASE) (DIPHENOL OXIDASE).
GN LCC4.
OS THANATEPHORUS CUCUMERIS (BLACK SCURF OF POTATO) (RHIZOCTONIA SOLANI).
OC EUKARYOTA; FUNGI; BASIDIOMYCOTA; HYMNOMYCETES; CERATOBASIDIALES;
OC CERATOBASIDIACEAE; THANATEPHORUS.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN-RS22:
RX MEDLINE: 96171523.
RA WAHLSTEINER J.A., XU F., BROWN K.M., BROWN S.H., GOLIGHTLY E.J.,
RA HALKIER T., KAUPPINEN S., PEDERSON A., SCHNEIDER P.;
RT "The identification and characterization of four laccases from the
RT plant pathogenic fungus Rhizoctonia solani.";
RL CURR. GENET. 29:395-403(1996).
CC -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLE). THIS ISOZYME SHOWS OPTIMAL ACTIVITY AT PH 7.
CC -!- CATALYTIC ACTIVITY: 4 BENZENEDIOL + O(2) -> 4 BENZOSEMIQUINONE +
CC 2 H(2)O.
CC -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC CENTRES KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- TISSUE SPECIFICITY: IN MYCELIA, AT A HIGHER LEVEL THAN LCC1, LCC2
CC AND LCC3.
CC -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 254277; E206168; .
DR PROSITE: P500079; MULTICOPPER_OXIDASE1; FALSE_NEG.
DR PROSITE: P500080; MULTICOPPER_OXIDASE2; FALSE_NEG.
DR PFAM: PF00394; Cu-oxidase; 2.
KW OXIDOREDUCTASE; SIGNAL; COPPER; METAL-BINDING; GLYCOPROTEIN; REPEAT;
KW LIGNIN DEGRADATION; MULTIGENE FAMILY; POLYMORPHISM.
FT SIGNAL 1 19
FT CHAIN 20 531
FT DOMAIN 23 146
FT DOMAIN 136 315
FT DOMAIN 384 507
FT METAL 83 83
FT METAL 85 85
FT METAL 128 128
FT METAL 130 130
FT METAL 427 427
FT METAL 430 430
FT METAL 432 432
FT METAL 479 479
FT METAL 480 480
FT METAL 481 481
FT METAL 487 487
FT METAL 7 ?
FT CARBOHYD 66 66
FT CARBOHYD 109 109
FT CARBOHYD 186 186
FT CARBOHYD 231 231
FT CARBOHYD 280 280
FT CARBOHYD 395 395
FT VARIANT 42 42
FT VARIANT 119 119
FT VARIANT 246 246
FT VARIANT 256 256
FT VARIANT 261 261
SQ SEQUENCE 531 AA: 57545 MW: C90CC591 CRC32:

Query Match 43.4%; Score 59; DB 1; Length 531;
Best Local Similarity 45.3%; Pred. No. 3.08e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 257 HOPLPVDSFQI 267
QY 338 QOPIPTETQV 348
:|:|:|:|:|:

RESULT 4
ID SAPA_SALTY STANDARD; PRT; 549 AA.
AC P36634;
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE PEPTIDE TRANSPORT PERIPLASMIC PROTEIN SAPA PRECURSOR.
GN SAPA.
OS SALMONELLA TYPHIMURIUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-140285;
RX MEDLINE: 94038887.
RA PARRA-LOPEZ C., BAER M.T., GROISMAN E.A.;
RT "Molecular genetic analysis of a locus required for resistance to
RT antimicrobial peptides in Salmonella typhimurium.";
RL EMBL J. 12:4053-4062(1993).
CC -!- FUNCTION: INVOLVED IN A PEPTIDE INTAKE TRANSPORT SYSTEM THAT
CC PLAYS A ROLE IN THE RESISTANCE TO ANTIMICROBIAL PEPTIDES.
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X74212; G414208; .
DR PIR: S39585; S39585
DR STYGENE; SG10377; SAPA.
DR PROSITE: P501040; SBP_BACTERIAL_5; 1.
DR PFAM: PF00496; SBP_bac_5; 1.
DR HSP: P23847; 1DPP.
KW PEPTIDE TRANSPORT; TRANSPORT; PERIPLASMIC; SIGNAL.
FT SIGNAL 1 21
FT CHAIN 22 549
FT PEPTIDE TRANSPORT PERIPLASMIC PROTEIN
SQ SEQUENCE 549 AA: 61622 MW: CE68A760 CRC32:

Query Match 43.4%; Score 59; DB 1; Length 549;
Best Local Similarity 43.8%; Pred. No. 3.08e+00;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 209 LLDROPVGTGPFQLSE 224
QY 335 LPDQOPIPTETQVAD 350
:|:|:|:|:|:

RESULT 5

ID RRPO_TNVA STANDARD; PRT; 723 AA.
AC P22958;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) [CONTAINS: 23 KD PROTEIN].
OS TOBACCO NECROSIS VIRUS (STRAIN A) (TNV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NECROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90320143.
RA MEULEWAETER F., SEURINCK J., VAN EMMELO J.;
RT "Genome structure of tobacco necrosis virus strain A.";
RL VIROLOGY 177:699-709(1990).
CC -!- READTHROUGH OF A TERMINATOR CODON OCCURS BETWEEN CODONS FOR
CC LYS-202 AND GLY-203.
CC -----

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M33002; G310915; -
 DR PIR; A35523; RRMOTN
 KW TRANSFERASE; RNA-DIRECTED RNA POLYMERASE.
 FT CHAIN 1 202 23 KD PROTEIN.
 FT VARIANT 72 72 V -> A.
 FT VARIANT 698 698 K -> R.
 SQ SEQUENCE 723 AA; 82167 MW; 62C87D02 CRC32;

Query Match 43.4%; Score 59; DB 1; Length 723;

Best Local Similarity 58.3%; Pred. No. 3.08e+00;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 705 IPTDTYQVSGEH 716

Y 341 IPTETQVADRH 352

RESULT 6

ID MEI3_SCHPO STANDARD; PRT; 148 AA.

AC P08090;

DT 01-AUG-1988 (REL. 08, CREATED)

DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE 21 KD PROTEIN INDUCING MEIOSIS AND SPORULATION.

GN MEI3 OR SPEC119.04.

OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;

OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;

CC SCHIZOSACCHAROMYCES.

CC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 87218534.

RA MCLEOD M., STEIN M., BEACH D.;

RT "The product of the mei3+ gene, expressed under control of the
 RT mating-type locus, induces meiosis and sporulation in fission
 RT yeast.";

RL EMBO J. 6:729-736(1987).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-972;

RA WOOD V., RAJANDREAM M.A., BARRELL B.G., SKELTON J., CHURCHER C.M.;

RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -!- FUNCTION: THIS PROTEIN, EXPRESSED UNDER CONTROL OF THE MATING-TYPE

CC LOCUS, INDUCES MEIOSIS AND SPORULATION IN FISSION YEAST. IT DOWN

CC REGULATES THE RAN1/PAT1 GENE PRODUCT.

CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X05142; G4993; -

DR EMBL; AL022117; E1263962; -

DR PIR; S07196; S07196.

KW MEIOSIS; SPORULATION.

SQ SEQUENCE 148 AA; 16473 MW; A1D53AE8 CRC32;

Query Match

Best Local Similarity 42.6%; Score 58; DB 1; Length 148;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 131 PRPNETFIDTR 142

Oy 340 PIPTEFQVADR 351
 | | | | | : : : | |

RESULT 7

ID NUSA_TETH STANDARD; PRT; 387 AA.

AC P48514;

DT 01-FEB-1996 (REL. 33, CREATED)

DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE N-UTILIZATION SUBSTANCE PROTEIN A HOMOLOG.

GN NUSA.

OS THERMUS AQUATICUS (SUBSP. THERMOPHILUS).

OC BACTERIA; THERMUS/DEINOCOCCUS GROUP; THERMUS GROUP; THERMUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-HB8;

RA VORNLOCHER H., SPRINZL M.;

RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -!- FUNCTION: COULD PARTICIPATES IN BOTH THE TERMINATION AND

CC ANTITERMINATION OF TRANSCRIPTION (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE NUSA FAMILY.

CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z480013; G542364; -

DR PFAM; PF00013; KH-domain; 1.

DR PFAM; PF00575; SL; 1

KW TRANSCRIPTION TERMINATION.

SQ SEQUENCE 387 AA; 43939 MW; 07FDCADE CRC32;

Query Match

Best Local Similarity 42.6%; Score 58; DB 1; Length 387;

Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 172 LPRSEIPTKYHPGQR 188

Y 335 LPDQPIPTETFEQVADR 351

RESULT 8

ID PPAF_ARATH STANDARD; PRT; 469 AA.

AC Q38924; Q42349;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE IRON(III)-ZINC(II) PURPLE ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2)

DE (PAP).

GN PAPI.

OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).

OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;

OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;

OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.

CC [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV; LANDSBERG ERRECTA;

RA PATEL K.S., LOCKLESS S.W., MCKNIGHT T.D.;

RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE OF 138-268 FROM N.A.

RC STRAIN-CV. COLUMBIA;

RA COOKE R., LAUDIE M., RAYNAL M., DELSENY M.;

RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O -- AN

CC ALCOHOL + ORTHOPHOSPHATE.

CC -!- COFACTOR: IRON AND ZINC (BY SIMILARITY).

CC -!- SUBUNIT: HOMODIMER, DISULFIDE LINKED (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48448; G1218042; -;
DR EMBL; F20043; E225669; -;
DR PFAM; PF00041; fn3; 1;
DR HSSP; P80366; 3KBP.
KW HYDROLASE; IRON; ZINC; METAL-BINDING; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 28
FT CHAIN 29 469
FT METAL 168 197
FT METAL 197 197
FT METAL 200 200
FT METAL 234 234
FT ACT_SITE 235 235
FT METAL 319 319
FT METAL 356 356
FT METAL 358 358
FT METAL 378 378
FT DISULFID 378 378
FT CARBOHYD 114 114
FT CARBOHYD 176 176
FT CARBOHYD 307 307
FT CARBOHYD 429 429
FT CONFLICT 151 151
FT CONFLICT 216 216
FT CONFLICT 264 264
SQ SEQUENCE 469 AA; 54138 MW; 60D9D146 CRC32;

Query Match 42.6%; Score 58; DB 1; Length 469;
Best Local Similarity 40.0%; Pred. No. 4.70e+00;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 40 LPDQPLDSDVFEVP 54
|||:|:|:|:|:
QY 335 LPDQQPIPTETFFQVA 349

RESULT 9
ID Y029_HUMAN STANDARD; PRT; 971 AA.
AC Q15032;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE HYPOTHETICAL PROTEIN KIAA0029.
GN KIAA0029.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE; 96051387.
RA NOMURA N., MIYAJIMA N., SAZUKA T., TANAKA A., KAWARABAYASI Y.,
RA SATO S., NAGASE T., SEKI N., ISHIKAWA K.-I., TABATA S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1".
RL DNA RES. 1:27-35(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D21852; E316530; -;
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 461 480 POLY-PRO.
FT DOMAIN 694 698 POLY-PRO.
SQ SEQUENCE 971 AA; 107673 MW; 735F538D CRC32;

Query Match 42.6%; Score 58; DB 1; Length 971;
Best Local Similarity 42.1%; Pred. No. 4.70e+00;
Matches 8; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Db 478 LPQGPVPTAGYPASG-HP 495
|||:|:|:|:|:
QY 335 LPDQQPIPTETFFQVADRH 353

RESULT 10
ID HOG3_HORVU STANDARD; PRT; 289 AA.
AC P80198;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DE GAMMA-HORDEIN 3.
DE GAMMA-HORDEIN 3.
OS HORDEUM VULGARE (BARLEY).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; HORDEUM.
RN [1]
RP SEQUENCE OF 1-11, AND SEQUENCE OF 4-289 FROM N.A.
RC STRAIN=CV. HOR2CA; TISSUE=ENDOSPERM;
RX MEDLINE; 94100996.
RA RECHINGER K.B., SIMPSON D.J., SVENDSEN I., CAMERON-MILLS V.;
RT "A role for gamma 3 hordein in the transport and targeting of
RT prolamin polypeptides to the vacuole of developing barley
RT endosperm".
RL PLANT J. 4:841-853(1993).
CC -!- FUNCTION: HAS A ROLE IN THE TRANSPORT AND TARGETING OF PROLAMINS
CC TO THE VACUOLE OF DEVELOPING BARLEY ENDOSPERM.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (AS GLOBULES) AND VACUOLAR
CC (AS PROTEIN BODIES).
CC -!- TISSUE SPECIFICITY: DEVELOPING ENDOSPERM.
CC -!- DOMAIN: SULFUR RICH HORDEIN WHICH POSSES AN N-TERMINAL HALF
CC COMPOSED OF PROLINE-GLUTAMINE BLOCKS ORGANIZED IN REPEATING UNITS
CC AND A C-TERMINAL HALF WHERE THE REPEATS ARE DISPERSED AND LESS
CC CONSERVED.
CC -!- SIMILARITY: TO GAMMA-GLIADIN FROM WHEAT, AND A GAMMA-SECALIN FROM
CC RYE, AND LESS TO A B1 HORDEIN FROM BARLEY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X72628; G288709; -;
DR PFAM; PF00234; tryp_alpha_aml1; 1
KW SEED STORAGE PROTEIN; MULTIGENE FAMILY.
SQ SEQUENCE 289 AA; 33189 MW; 5345ADF CRC32;

Query Match 41.9%; Score 57; DB 1; Length 289;
Best Local Similarity 44.4%; Pred. No. 7.14e+00;
Matches 8; Conservative 6; Mismatches 3; Indels 1; Gaps 1;

Db 49 LPQQPFPQOP-QLPHOH 65
|||:|:|:|:|:
QY 335 LPDQQPIPTETFFQVADRH 352

RESULT 11

```
ID YD3_SCHPO STANDARD; PRT; 548 AA.
AC Q10428;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 63.3 KD PROTEIN C189.03 IN CHROMOSOME I.
GN SPAC189.03.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA CONNOR R.; CHURCHER C.M.; BARRELL B.G.; RAJANDREAM M.A.; WALSH S.V.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: STRONG, TO YEAST RTS1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X66930; G9144; -
CC FLYBASE; FBgn0012901; Dslm\tra.
CC DR FLYBASE; FBgn0012901; Dslm\tra.
CC DR NUCLEAR PROTEIN; ALTERNATIVE SPLICING; SEXUAL DIFFERENTIATION.
CC FT DOMAIN 15 112
CC KW NUCLEAR PROTEIN; ARG/SER-RICH (RS DOMAIN).
CC SQ SEQUENCE 184 AA; 22386 MW; F7F6B3E CRC32;
CC -----
Query Match 41.2%; Score 56; DB 1; Length 184;
Best Local Similarity 36.8%; Pred. No. 1.08e+01;
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
Db 149 LPRPPFPFPAPFYRORQP 167
QY || | | | | | | | |
QY 335 LPDQPIPTETQVADRHP 353
CC -----
RESULT 13
ID ETVI_MOUSE STANDARD; PRT; 477 AA.
AC P41164;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ER81 PROTEIN (ETS TRANSLLOCATION VARIANT 1).
GN ETVI OR ETSR81 OR ER81.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RA BROWN T.A.; MCKNIGHT S.L.;
RX MEDLINE; 94040714.
RT "Specificities of protein-protein and protein-DNA interaction of GABP
RT alpha and two newly defined ets-related proteins.";
RL GENES DEV. 6:2502-2512(1992).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO DNA SEQUENCES
CC CONTAINING THE CONSENSUS PENTANUCLEOTIDE 5'-CGGA[AT]-3'.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: ABUNDANT IN KIDNEY. MODERATE LEVELS SEEN IN
CC THE HEART, BRAIN, LUNG, EMBRYO AND LOWER LEVELS SEEN IN SPLEEN,
CC INTESTINE, TESTIS AND THYMUS.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L10426; G515964; -
CC FIR; B46396; B46396.
CC DR MGI; 99254; ETSR81.
CC DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
CC DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
CC DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
CC DR PFAM; PF00178; Ets; 1.
CC DR HSP; P14921; 2SW.
CC KW TRANSCRIPTION REGULATION; NUCLEAR PROTEIN; DNA-BINDING.
CC FT DNA_BIND 335 415 ETS-DOMAIN.
CC SQ SEQUENCE 477 AA; 55040 MW; 670ADC77 CRC32;
CC -----
ID TRSF_DROSI STANDARD; PRT; 184 AA.
AC Q24669;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE FEMALE-SPECIFIC TRANSFORMER PROTEIN.
GN TRA.
OS DROSOPHILIA SIMULANS (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92275311.
RA O'NEIL M.T.; BELOTE J.M.;
RT "Interspecific comparison of the transformer gene of Drosophila
RT reveals an unusually high degree of evolutionary divergence.";
RL GENETICS 131:113-128(1992).
CC -1- FUNCTION: REGULATES FEMALE SOMATIC SEXUAL DIFFERENTIATION.
CC ACTIVATES DSX FEMALE-SPECIFIC SPLICING BY PROMOTING THE FORMATION
CC OF A SPLICING ENHANCER COMPLEX WHICH CONSISTS OF TRA, TRA2 AND SR
CC PROTEINS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; SPECKLED SUBNUCLEAR COMPARTMENT (BY
CC SIMILARITY).
CC -1- DOMAIN: RS DOMAIN DIRECTS LOCALIZATION OF PROTEINS TO THE SPECKLED
CC SUBNUCLEAR COMPARTMENT AND THE PURPOSE OF THIS LOCALIZATION IS TO
CC ALLOW COLOCALIZATION AND CO-CONCENTRATION OF COMPONENTS OF THE
CC SPLICING AND SPLICING REGULATORY MACHINERY TO PERMIT RELATIVELY
CC HIGH RATES AND/OR EFFICIENCIES OF REACTION AND INTERACTION (BY
CC SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: THE SEXUAL REGULATION OF TRA OCCURS THROUGH
CC A MECHANISM OF SEX-SPECIFIC ALTERNATIVE RNA SPLICING. FEMALE-
CC SPECIFIC CDNAS ARE SHORTER THAN THE NON-SEX-SPECIFIC CDNAS. THE
```

Query Match 41.2%; Score 56; DB 1; Length 477;
Best Local Similarity 41.2%; Pred. No. 1.08e+01;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 169 LPPQSIPDSTYPMOHR 185
|||:|:|:|:|:|:|
QY 335 LPDQOPIPTETFOADR 351

RESULT 14
ID ERK3 RAT STANDARD; PRT; 720 AA.
AC P27704;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE EXTRACELLULAR SIGNAL-REGULATED KINASE 3 (EC 2.7.1.1-) (ERK3) (P55-MAPK)
DE MAPK)
GN PRK5 OR ERK3
OS RATTUS NORVEGICUS (RAT)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN-SPRAGUE-DAWLEY; TISSUE=BRAIN, AND LIVER;
RX MEDLINE; 91235302.
RA BOULTON T.G., NYE S.H., ROBBINS D.J., IP N.Y., RADZIEJEWSKA E.,
RA MORGENESSER S.D., DEPINHO R.A., PANAYOTATOS N., COBB M.H.,
RA YANCOPOULOS G.D.;
RT "ERKs: a family of protein-serine/threonine kinases that are
activated and tyrosine phosphorylated in response to insulin and
RT NGF.";
RL CELL 65:663-675(1991).
RN [2]
RP IDENTIFICATION OF PROBABLE FRAMESHIFT.
RA HULO-DEMOLE C., BRACONI-QUINTAJE S.;
RL UNPUBLISHED OBSERVATIONS (MAR-1997).
CC -!- FUNCTION: PHOSPHORYLATES MICROTUBULE-ASSOCIATED PROTEIN-2 (MAP2).
CC -!- MAY PROMOTES ENTRY IN THE CELL CYCLE.
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS WITHIN THE NERVOUS SYSTEM,
EXPRESSED IN DIFFERENT TISSUES, MOSTLY IN SKELETAL MUSCLE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT HIGHEST LEVELS EARLY IN
DEVELOPMENT.
CC -!- SIMILARITY: BELONGS TO THE CDC2/CDC28 SUBFAMILY OF SER/THR
PROTEIN KINASES. STRONGEST SIMILARITY WITH OTHER MAP KINASES.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
INTRODUCED IN POSITION 503 TO EXTEND THE SIMILARITY WITH THE HUMAN
HOMOLOG.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; M64301; G204058; ALT_FRAME.
DR PIR; B40033; B40033.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; P27703; IGL0.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING; CELL CYCLE.
FT DOMAIN 20 316 PROTEIN KINASE.
FT NP_BIND 26 34 ATP (BY SIMILARITY).
FT BINDING 49 49 ATP (BY SIMILARITY).
FT ACT_SITE 152 152 BY SIMILARITY.
SQ SEQUENCE 720 AA; 82274 MW; 93460A24 CRC32;

Query Match 41.2%; Score 56; DB 1; Length 720;
Best Local Similarity 31.3%; Pred. No. 1.08e+01;

Matches 5; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Db 321 PPTDEPISSHPHIED 336
:|:|:|:|:|:|:|:|:|
QY 335 LPDQOPIPTETFOADR 350

RESULT 15
ID EBN3_EBV STANDARD; PRT; 812 AA.
AC P12977; Q66540;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE EBNA-3 NUCLEAR PROTEIN (EBNA-3A).
DE BERF1.
GN EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).
OS VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE; LYMPHOCRYPTOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84270667.
RA BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
RA GIBSON T.J., HATFULL G., HUDSON G.S., SATCHELL S.C., SEGUIN C.,
RA TUFENELL P.S., BARRELL B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL NATURE 310:207-211(1984).
RN [2]
RP CHARACTERIZATION.
RX HENNESSY K., WANG F., BUSHMAN E.W., KIEFF E.;
RX MEDLINE; 86287322.
RT "Definitive identification of a member of the Epstein-Barr virus
nuclear protein 3 family.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:5693-5697(1986).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE; 90266473.
RA PETTI L., SAMPLE C., KIEFF E.;
RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
latent infection nuclear proteins.";
RL VIROLOGY 176:563-574(1990).
CC -!- FUNCTION: INVOLVED IN LATENT CYCLE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH THE NUCLEAR MATRIX.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; V01555; E25053; ALT_INIT.
KW NUCLEAR PROTEIN.
SQ SEQUENCE 812 AA; 89097 MW; B58F8DC6 CRC32;

Query Match 41.2%; Score 56; DB 1; Length 812;
Best Local Similarity 52.9%; Pred. No. 1.08e+01;
Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Db 504 VPEQOMFPGAPFSQVAD 520
||:|:|:|:|:|:|:|:|:|
QY 335 LPDQOPIPTETFOADR 350

Search completed: Thu Jul 8 18:46:43 1999
Job time : 11 secs.

THIS PAGE BLANK (USPTO)

W P S R E L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:47:01 1999; MasPar time 5.59 Seconds
Tabular output not generated. 185.411 Million cell updates/sec

Title: >US-09-041-236-2
Description: (335-353) from US09041236.pep (20 of 45)
Perfect Score: 136
Sequence: 1 LPDQOPIPTETFOVADHRP 19

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 27.150; Variance 38.018; scale 0.714

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	136	100.0	666	4	075326 SEMAPHORIN L.	2.55e-17
2	112	82.4	393	11	088371 SEMAPHORIN L (FRAGMENT)	2.09e-11
3	95	69.9	653	14	064906 SIMILAR TO GENBANK ACC	2.00e-07
4	72	52.9	299	2	034973 YTPA.	1.87e-02
5	65	47.8	1147	2	P94448 PYRUVATE CARBOXYLASE (4.48e-01
6	64	47.1	934	5	044164 COSMID F16B3.	6.94e-01
7	63	46.3	665	3	000611 KINASE.	1.07e+00
8	63	46.3	1124	5	045401 F23B2.4 PROTEIN.	1.07e+00
9	61	44.9	337	2	066618 TRANSCRIPTION TERMINAT	2.52e+00
10	61	44.9	990	5	046086 1-EVIDENCE-PREDICTED B	2.52e+00
11	61	44.9	1043	5	023999 PUTATIVE SINGLE-STRAND	2.52e+00
12	61	44.9	2109	14	065667 237K PROTEIN.	2.52e+00
13	61	44.9	2109	14	083743 237K ORF.	2.52e+00
14	60	44.1	155	6	Q28051 AMELOGENIN (FRAGMENT).	3.84e+00
15	60	44.1	250	2	G50315 ORF250.	3.84e+00
16	60	44.1	1458	5	Q08095 DYNEIN HEAVY CHAIN, CY	5.83e+00
17	59	43.4	1458	5	Q08095 DYNEIN HEAVY CHAIN, CY	5.83e+00
18	58	42.6	170	14	070642 REX PROTEIN.	8.79e+00
19	58	42.6	347	10	Q40055 C HORDEIN PRECURSOR.	8.79e+00
20	58	42.6	379	10	Q64550 YUP8H2R.17 PROTEIN.	8.79e+00

21	58	42.6	541	5	Q20214 F40F12.4 PROTEIN.	8.79e+00
22	58	42.6	610	2	P74396 HYPOTHETICAL 65.4 KD P	8.79e+00
23	58	42.6	619	2	Q82912 NIKB.	8.79e+00
24	57	41.9	408	10	Q39532 OPAQUE 2.	1.32e+01
25	57	41.9	1031	10	Q22575 GLYCINE DECARBOXYLASE	1.32e+01
26	57	41.9	1032	10	Q38766 VICTORIN BINDING PROTE	1.32e+01
27	57	41.9	1551	2	Q51579 FERREDOXIN-DEPENDENT G	1.32e+01
28	57	41.9	2150	5	Q44131 C08G9.2 PROTEIN.	1.32e+01
29	56	41.2	206	2	Q05148 PUTATIVE REGULATORY PR	1.97e+01
30	56	41.2	260	14	Q67647 INFECTIOUS LARYNGOTRAC	1.97e+01
31	56	41.2	267	10	Q24353 MEN-5 PROTEIN.	1.97e+01
32	56	41.2	296	2	Q85288 ASPARTIC ACID AMINASE.	1.97e+01
33	56	41.2	405	2	P77991 ASPARTYLKINASE ALPHA-2	1.97e+01
34	56	41.2	405	2	P97151 ASPARTATE KINASE ALPHA	1.97e+01
35	56	41.2	457	13	Q90269 INTRACELLULAR TYROSINE	1.97e+01
36	56	41.2	499	5	Q21295 K07F5.12 PROTEIN.	1.97e+01
37	56	41.2	907	14	Q69021 MRNA T2 OF 84 KB TRANS	1.97e+01
38	56	41.2	1508	5	Q46014 ZK337.1A PROTEIN.	1.97e+01
39	56	41.2	1519	5	Q46015 ZK337.1B PROTEIN.	1.97e+01
40	56	41.2	1539	4	Q15023 KIAA0305.	1.97e+01
41	55	40.4	616	4	Q75827 DJ71L16.5 (K1AA0267 LI	2.93e+01
42	55	40.4	699	4	Q43263 RNA EDITING DEAMINASE	2.93e+01
43	55	40.4	972	14	Q82720 POLYPROTEIN.	2.93e+01
44	55	40.4	972	14	Q82731 POLYPROTEIN.	2.93e+01
45	55	40.4	972	14	P89521 POLYPROTEIN.	2.93e+01

ALIGNMENTS

RESULT 1					
ID 075326	PRELIMINARY;	PRT;	666	AA.	
AC 075326					
DT 01-NOV-1998	(TREMBLREL. 08, CREATED)				
DT 01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)				
DT 01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)				
DE SEMAPHORIN L.					
GN SEMAL.					
OS HOMO SAPIENS (HUMAN).					
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;					
OC CATARRHINI; HOMINIDAE; HOMO.					
RN [1]					
RP SEQUENCE FROM N.A.					
RX MEDLINE; 98389619.					
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;					
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA					
RT viruses."					
RL GENOMICS 51:340-350(1998).					
DR EMBL; AF030698; G3523115; ..					
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;					

Query Match 100.0%; Score 136; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 2.55e-17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 367	LPDQOPIPTETFOVADHRP	385
QY 335	LPDQOPIPTETFOVADHRP	353

RESULT 2
ID 088371
AC 088371;

DT 01-NOV-1998	(TREMBLREL. 08, CREATED)	
DT 01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)	
DT 01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)	
DE SEMAPHORIN L (FRAGMENT).		
GN SEMAL.		
OS MUS MUSCULUS (MOUSE).		
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.		
RN [1]		
RP SEQUENCE FROM N.A.		
RX MEDLINE; 98389619.		

```

RA  LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT  "New eukaryotic semaphorins with close homology to semaphorins of DNA
RL  viruses.";
RL  GENOMICS 51:340-350(1998).
FT  EMBL; AF030699; G3523117; -.
DR  NON_TER 393 393
SQ  SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;

Query Match      82.4%; Score 112; DB 11; Length 393;
Best Local Similarity 84.2%; Pred. No. 2.09e-11;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db  365 LPKQPIPTTFQVADSRP 383
    || ||||| ||||| ||
QY  335 LPDQPIPTTFQVADSRP 353

RESULT 3
ID  Q64906 PRELIMINARY; PRT; 653 AA.
AC  Q64906;
DT  01-NOV-1996 (TREMBLREL. 01, CREATED)
DT  01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE  SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS  ALCELAHINE HERPESVIRUS 1.
OC  VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC  GAMMAHERPESVIRINAE.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C500;
RX  MEDLINE; 97201573.
RA  ENSSER A., FLECKENSTEIN B.;
RT  "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL  J. GEN. VIROL. 76:1063-1067(1995).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C500;
RX  MEDLINE; 97404659.
RA  ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RT  "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL  J. VIROL. 71:6517-6525(1997).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C500;
RL  ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RL  SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL; U18243; G1000717; -.
DR  EMBL; AF005370; G2337970; -.
SQ  SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match      69.9%; Score 95; DB 14; Length 653;
Best Local Similarity 63.2%; Pred. No. 2.00e-07;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db  389 MKNHQVPTTFQVADSRP 407
    : : : ||||| ||||| |
QY  335 LPDQPIPTTFQVADSRP 353

RESULT 4
ID  Q34973 PRELIMINARY; PRT; 299 AA.
AC  Q34973;
DT  01-JAN-1998 (TREMBLREL. 05, CREATED)
DT  01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT  01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE  YTAIP.
GN  BACILLUS SUBTILIS.
OS  BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC  BACILLUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 96312354.

```

```

RA  BOWER S., PERKINS J.B., YOCUM R.R., HOWITT C.L., RAHAIM P., PERO J.;
RT  "Cloning, sequencing, and characterization of the Bacillus subtilis
RT  biotin biosynthetic operon.";
RL  J. BACTERIOL. 178:4122-4130(1996).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  LAPIDUS A., GALLERON N., SOROKIN A., EHRLICH D.;
RN  SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=168;
RX  MEDLINE; 98044033.
RA  KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALONI G.,
RA  AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA  BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA  BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA  CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
RA  DENIZOT F., DEVINE K.M., DUSTERHOFF A., EHRLICH S.D., EMERSON P.T.,
RA  ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA  FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA  GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA  GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA  HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA  JORIS B., KARAWATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA  KOBAYASHI Y., KOETTER P., KONIGSTEIN G., KROGH S., KUMANO M.,
RA  KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA  LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA  MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA  NOONE D., O'REILLY M., OGAWA K., OGAWARA A., OUDEGA B., PARK S.H.,
RA  PARRO V., POHL T.M., PORTELELLA D., FORWOLLEK S., PRESCOTT A.M.,
RA  PRESECAN E., PUJIC P., PURNELLE B., RAPPOPORT G., REY M., REYNOLDS S.,
RA  RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA  SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA  SEKIGUCHI J., SEKOWSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDI B.,
RA  SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA  TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA  TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA  VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA  WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA  YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT  "The complete genome sequence of the gram-positive bacterium Bacillus
RT  subtilis.";
RL  NATURE 390:249-256(1997).
RN  [4]
RP  SEQUENCE FROM N.A.
RC  STRAIN=168;
RA  KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL  SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL; AF008220; G2293182; -.
DR  EMBL; Z99119; E1185898; -.
SQ  SEQUENCE 299 AA; 33776 MW; EB5CA92D CRC32;

Query Match      52.9%; Score 72; DB 2; Length 299;
Best Local Similarity 47.1%; Pred. No. 1.87e-02;
Matches 8; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db  16 LPDRRPISVETLRIBER 32
    |||:|:|:|:|:|:|
QY  335 LPDQPIPTTFQVADR 351

RESULT 5
ID  P94448 PRELIMINARY; PRT; 1147 AA.
AC  P94448;
DT  01-MAY-1997 (TREMBLREL. 03, CREATED)
DT  01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DE  PYRUVATE CARBOXYLASE (EC 6.4.1.1).
OS  BACILLUS STEROOTHERMOPHILUS.
OC  BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
RN  [1]
RP  SEQUENCE FROM N.A.

```


ID	Q00611	PRELIMINARY;	PRT;	665 AA.
AC	Q00611;			
DT	01-NOV-1996	(TREMBLREL. 01, CREATED)		
DT	01-NOV-1996	(TREMBLREL. 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	KINASE.			
GN	TB3.			
OS	COLLETOTRICHUM TRIFOLI.			
OC	EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;			
OC	PHYLLACHORALES; PHYLLACHORACEAE; MITOSPORIC PHYLLACHORACEAE;			
OC	COLLETOTRICHUM.			
[1]				
RN	SEQUENCE FROM N.A.			
RP	MEDLINE; 96323031.			
RX	BUHR T.L., TRUESDELL G.M., DICKMAN M.B., OVED S., HUANG C.,			
RA	YARDEN O.;			
RT	"A kinase-encoding gene from Colletotrichum trifolii complements a			
RT	colonial growth mutant of Neurospora crassa."			
RI	MOL. GEN. GENET. 251:565-572(1996).			
DR	ENBL; U14989; G641979.			
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.			
DR	PFAM; PF00089; pkinase; 1.			
DR	SEQUENCE 665 AA; 76015 MW; 88C4543D CRC32;			
QY				
Db	183 LPNOQPASVDPFPAPAPERN 201			
QY	335 LPDQOPIPTETF-QVADRH 352			
RESULT	8			
ID	Q45401	PRELIMINARY;	PRT;	1124 AA.
AC	Q45401;			
DT	01-JUN-1998	(TREMBLREL. 06, CREATED)		
DT	01-JUN-1998	(TREMBLREL. 06, LAST SEQUENCE UPDATE)		
DT	01-JAN-1999	(TREMBLREL. 09, LAST ANNOTATION UPDATE)		
DE	F33B2.4 PROTEIN.			
GN	F33B2.4.			
OS	CAENORHABDITIS ELEGANS.			
OC	EUKARYOTA; METAZOA; NEMATODA; SCERENTEA; RHABDITIDA; RHABDITIDA;			
OC	RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MCURRAY A.;			
RL	SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.			
[2]				
RN	SEQUENCE FROM N.A.			
RP	MEDLINE; 94150718.			
RA	WILSON R., AINCROUGH R., ANDERSON K., BAYNES C., BERKS M.,			
RA	BRONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,			
RA	CRAFTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,			
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JETER M., JOHNSTON L.,			
RA	JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,			
RA	LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,			
RA	PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,			
RA	SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,			
RA	THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,			
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RT	elegans";			
RL	NATURE 368:32-38(1994).			
DR	ENBL; Z82266; E1345942.			
QY	SEQUENCE 1124 AA; 127879 MW; 90DAC9B9 CRC32;			
Query Match	46.3%;	Score 63;	DB 5;	Length 1124;
Best Local Similarity	63.6%;	Pred. No. 1.07e+00;		
Matches	7;	Conservative 3;	Mismatches 1;	Indels 0; Gaps 0;
Db	884 TDAFAIADRP 894			

QY 343 TETFQVADHRP 353

RESULT 9
ID O66618 PRELIMINARY; PRT: 337 AA.
AC O66618;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION TERMINATION NUSA.
GN NUSA.
OS AQUIFEX ACOLICUS.
OC BACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT acolicus";
RL NATURE 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AE000681; G2982967;
SQ SEQUENCE 337 AA; 38494 MW; 994AC763 CRC32;

Query Match 44.9%; Score 61; DB 2; Length 337;
Best Local Similarity 35.3%; Pred. No. 2.52e+00;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 139 LPKREOLPTDNEFTGDK 155
|||:||||:|:
QY 335 LPDQOPIPTETFOVADR 351

RESULT 10
ID O46086 PRELIMINARY; PRT: 990 AA.
AC O46086;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE 1-EVIDENCE-PREDICTED BY CONTENT.
GN EG:63B12.6.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RA FERRAZ C., VIDAL S., BRUN C., BUCHETON A., DEMAILLIE J.G.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA BENOS P.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AL021106; E1226212;
SQ SEQUENCE 990 AA; 109699 MW; 5783D49D CRC32;

Query Match 44.9%; Score 61; DB 5; Length 990;
Best Local Similarity 33.3%; Pred. No. 2.52e+00;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 103 EQEPLPPDSMAYSER 117
:|:|:|:|:|:
QY 337 DOOPIPTETFOVADR 351

RESULT 11
ID Q23999 PRELIMINARY; PRT: 1043 AA.
AC Q23999;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TART-B1;
RX MEDLINE; 95107987.
RA SHEEN F., LEVIS R.W.;
RT "Transposition of the LINE-like retrotransposon TART to Drosophila
RT chromosome termini";
RL PROC. NATL. ACAD. SCI. U.S.A. 91:12510-12514(1994).
DR EMBL; U14101; G503663;
DR FLYBASE; FB00004904; TART-element.
SQ SEQUENCE 1043 AA; 115121 MW; 1BF4F70A CRC32;

Query Match 44.9%; Score 61; DB 5; Length 1043;
Best Local Similarity 38.9%; Pred. No. 2.52e+00;
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Db 699 LHSQHPKSAKGFQIVIRH 716
|:|:|:|:
QY 335 LPDQOPIPTETFOVADR 352

RESULT 12
ID Q65667 PRELIMINARY; PRT: 2109 AA.
AC Q65667;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE 237K PROTEIN.
OS BEET NECROTIC YELLOW VEIN VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FUROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S;
RX MEDLINE; 97128991.
RA SATO M., KIGUCHI T., KUSUME T., TAMADA T.;
RT "Complete nucleotide sequence of the Japanese isolate S of beet
RT necrotic yellow vein virus and comparison with European isolates";
RL ARCH. VIROL. 141:2163-2175(1996).
DR EMBL; D84410; D1013017;
DR PFAM; PF00978; RNA_dep_RNAPol2; 1.
SQ SEQUENCE 2109 AA; 237517 MW; 86C5DE7F CRC32;

Query Match 44.9%; Score 61; DB 14; Length 2109;
Best Local Similarity 50.0%; Pred. No. 2.52e+00;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 923 HOKIPTQAFESVR 936
:|:|:|:|:|:
QY 338 QOPIPTETFOVADR 351

RESULT 13
ID Q83743 PRELIMINARY; PRT: 2109 AA.
AC Q83743;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 237K ORF.
OS BEET NECROTIC YELLOW VEIN MOSAIC VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOBAMOVIRUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE F2(FOR CLONE PBF5), ISOLATE F13(OTHER CL.);
RA BOUZOBAA S., QUILLET L., GUILLET H., JONARD G., RICHARDS K.;
RL J. GEN. VIROL. 88:615-626(1987).
DR EMBL; X05147; G60643; -
DR EMBL; D00115; D1000513; -
DR PFAM; PF00978; RNA_dep_RNApo12; 1.
SQ SEQUENCE 2109 AA; 237492 MW; C4D2BA2B CRC32;

Query Match 44.9%; Score 61; DB 14; Length 2109;
Best Local Similarity 50.0%; Pred. No. 2.52e+00;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 923 HOKIPTQAFYSVR 936
QY 338 QOPIPTETFOVADR 351
:|||||:|

RESULT 14
ID Q28051 PRELIMINARY; PRT; 155 AA.
AC Q28051;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE AMELOGENIN (FRAGMENT).
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87166009.
RA SHIMOKAWA H., SOBEL M.E., SASAKI M., TERMINE J.D., YOUNG M.F.;
RT "Heterogeneity of amelogenin mRNA in the bovine tooth germ.";
RL J. BIOL. CHEM. 262:4042-4047(1987).
DR EMBL; J02696; G162658; -
FT NON_TER 1
SQ SEQUENCE 155 AA; 17482 MW; 3E7EA17C CRC32;

Query Match 44.1%; Score 60; DB 6; Length 155;
Best Local Similarity 36.8%; Pred. No. 3.84e+00;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 67 LPAQOPQPSIQPHOP 85
QY 335 LPDQPIPTETFOVADRHP 353
|||||:|

RESULT 15
ID Q50315 PRELIMINARY; PRT; 250 AA.
AC Q50315;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ORF250.
OS MYCOPLASMA PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-M129;
RX MEDLINE; 96177562.
RA HILBERT H., HIMMELREICH R., PLAGENS H., HERRMANN R.;
RT "Sequence analysis of 56 kb from the genome of the bacterium
Mycoplasma pneumoniae comprising the dnaA region, the atp operon and
a cluster of ribosomal protein genes.";
RL NUCLEIC ACIDS RES. 24:628-639(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-M129;
RX MEDLINE; 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.C.,
RA HERRMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma

RT pneumoniae";
RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-M129;
RA HIMMELREICH R., HILBERT H., LI B.C.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U34816; G1209519; -
DR EMBL; AE000017; G1673816; -
SQ SEQUENCE 250 AA; 28662 MW; 96E26EB8 CRC32;
Query Match 44.1%; Score 60; DB 2; Length 250;
Best Local Similarity 41.2%; Pred. No. 3.84e+00;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 139 NQPVATSNIQVEATOP 155
QY 337 DQPIPTETFOVADRHP 353
:|||||:|

Search completed: Thu Jul 8 18:47:19 1999
Job time : 18 secs.

THIS PAGE BLANK (USPTO)

WQSRH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:50:46 1999; MasPar time 10.81 Seconds
Tabular output not generated. 19.672 Million cell updates/sec

Title: >US-09-041-236-2
Description: (355-364) from US09041236.pep (21 of 45)
Perfect Score: 74
Sequence: 1 VAQRVPMGP 10

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part3 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 16.495; Variance 45.633; scale 0.361

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	49	66.2	174	3	Recombinant CRIPTO CR	6.53e+01
2	49	66.2	174	26	Recombinant human CR	6.53e+01
3	49	66.2	188	39	Human CRIPTO-related	6.53e+01
4	49	66.2	188	34	Homo sapiens CRIPTO p	6.53e+01
5	49	66.2	188	4	Human CRIPTO CR-1 pro	6.53e+01
6	49	66.2	188	26	Human CRIPTO protein.	6.53e+01
7	49	66.2	589	6	Sequence encoded by t	6.53e+01
8	49	66.2	589	2	Polyhydroxybutyrate p	6.53e+01
9	49	66.2	672	13	Poly-beta-hydroxyvalka	6.53e+01
10	49	66.2	763	6	Gastrin-binding prote	6.53e+01
11	47	63.5	74	29	Human matrix metallo	1.11e+02
12	47	63.5	74	29	Alpha-v-beta-5 antago	1.11e+02
13	47	63.5	108	29	Alpha-v-beta-5 antago	1.11e+02
14	47	63.5	108	29	Human matrix metallo	1.11e+02
15	47	63.5	191	23	Retroviral reverse tr	1.11e+02
16	47	63.5	193	29	Alpha-v-beta-5 antago	1.11e+02

17	47	63.5	193	29	W41084	Human matrix metallo	1.11e+02
18	47	63.5	222	29	W41083	Human matrix metallo	1.11e+02
19	47	63.5	222	29	W41228	Alpha-v-beta-5 antago	1.11e+02
20	47	63.5	429	29	W41112	Human matrix metallo	1.11e+02
21	47	63.5	631	2	R07969	Complete type IV coll	1.11e+02
22	47	63.5	631	2	P96143	Sequence of human ty	1.11e+02
23	47	63.5	631	29	W41226	Human mature matrix m	1.11e+02
24	47	63.5	631	1	P91139	Human type IV collage	1.11e+02
25	47	63.5	660	1	R06420	Type IV collagenase c	1.11e+02
26	46	62.2	188	39	W87631	Human CRIPTO-related	1.44e+02
27	46	62.2	188	23	W19980	Human CRIPTO-related	1.44e+02
28	46	62.2	188	24	W25667	Protein encoded by CR	1.44e+02
29	46	62.2	188	4	R22547	Human CRIPTO-related	1.44e+02
30	45	60.8	271	29	W42383	Thermococcus celer al	1.87e+02
31	45	60.8	326	32	W60112	Mycobacterium vaccae	1.87e+02
32	45	60.8	537	28	W44195	Mouse osteoclast tran	1.87e+02
33	45	60.8	1430	30	W49907	Human pancreatic carc	1.87e+02
34	45	60.8	1436	28	W41361	Receptor protein tyro	1.87e+02
35	44	59.5	447	1	P93037	Chimeric antibody hea	2.42e+02
36	44	59.5	468	3	R13061	Monoclonal antibody O	2.42e+02
37	44	59.5	469	8	R40384	Monoclonal antibody M	2.42e+02
38	44	59.5	477	9	R47450	T84.12 Heavy chain.	2.42e+02
39	44	59.5	1007	33	W61246	Streptococcus pneumon	2.42e+02
40	43	58.1	210	9	R46555	fbp1, for reverse gen	3.13e+02
41	43	58.1	253	22	W20740	H. pylori cytoplasmic	3.13e+02
42	43	58.1	257	29	W42387	Thermococcus Cl-2 alk	3.13e+02
43	43	58.1	276	18	W00104	Pimelyl CoA synthase.	3.13e+02
44	43	58.1	341	2	R07684	CBEVP spheroidin.	3.13e+02
45	43	58.1	925	30	W55729	H. pylori ORF 05ep108	3.13e+02

ALIGNMENTS

RESULT 1
ID R13326 standard; protein; 174 AA.
AC R13326;
DT 22-OCT-1991 (first entry)
DE Recombinant CRIPTO protein.
KW Cancer.
OS Homo sapiens.
PN US7530165-A.
PD 09-JUL-1991.
PF 29-MAY-1990; 143529.
PR 29-MAY-1990; US-530165.
FA (USSH) NAT INST OF HEALTH.
PI Persico M, Solomon D;
DR WPI; 91-245692/33.
PT New human CRIPTO gene - used for developing prods. for the study,
PT diagnosis, prognosis and treatment of human cancers.
PS Disclosure; Fig 2; 24pp; English.
CC The sequence given is that of the recombinant E.coli derived
CC CRIPTO protein, produced from CRIPTO cDNA. The availability of
CC this protein make it possible to detect cells expressing the CRIPTO
CC gene. It can be used to develop oligonucleotide probes, antisense
CC oligonucleotides and antibodies for the study, diagnosis, prognosis
CC and treatment of human cancer. The cDNA sequence and the amino acid
CC sequence of the corresponding natural CRIPTO protein (different
CC to the recombinant sequence) are given in the specification but
CC poor print quality prevents their inclusion in this database.
SQ Sequence 174 AA;

Query Match 66.2%; Score 49; DB 3; Length 174;
Best Local Similarity 75.0%; Pred. No. 6.53e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 50 sqrvppmg 57

Qy 356 AQRVPMGP 363

RESULT 2
ID W32107 standard; Protein; 174 AA.

AC W32107;
DT 17-FEB-1998 (first entry)
DE Recombinant human CRIPTO protein expressed in E. coli.
KW CRIPTO gene; human; colorectal carcinoma; breast carcinoma; cancer;
KW tumour; diagnosis; prognosis; therapy; antibody; Escherichia coli.
OS Homo sapiens.
OS Synthetic.
PN US5654140-A.
PD 05-AUG-1997.
PF 29-MAY-1990; 530165.
PR 29-MAY-1990; US-530165.
PR 18-SEP-1992; US-947315.
PR 10-NOV-1994; US-337911.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Persico MG, Salomon DS;
DR WPI; 97-401840/37.
PT New cloned CRIPTO gene - used for diagnosis and prognosis of
PT carcinoma, also antisense sequences for tumour treatment
PS Disclosure; Fig 2; 13pp; English.
CC This protein comprises recombinant human CRIPTO produced in
CC Escherichia coli. Human CRIPTO cDNA (see T88973) was expressed
CC in E. coli inclusion bodies as a 20-22 kDa protein. CRIPTO is
CC a member of the epidermal growth factor family. Elevated levels
CC of CRIPTO mRNA, or of protein translated from it, are diagnostic
CC and prognostic of carcinoma, especially colorectal and breast
CC carcinoma. Recombinant CRIPTO can be expressed in prokaryotic or
CC eukaryotic host cells, and used for the identification and
CC characterisation of specific cell surface receptors, or to raise
CC specific antibodies to assay CRIPTO protein expression in
CC immunosays.
SO Sequence 174 AA;

Query Match 66.2%; Score 49; DB 26; Length 174;
Best Local Similarity 75.0%; Pred. No. 6.53e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 50 sqrvppmg 57
:|||||
QY 356 AQRVEPMG 363

RESULT 3
ID W87630 standard; Protein; 188 AA.
AC W87630;
DT 03-MAR-1999 (first entry)
DE Human CRIPTO-related (CR-1) protein.
KW human CRIPTO-related protein; CR-1; antibody; CR-3; proliferation;
KW differentiation; transformation; mesenchymal cell; epithelial cell;
KW tumour specific marker; cancer.
OS Homo sapiens.
PN US5854399-A.
PD 29-DEC-1998.
PF 05-JUN-1995; 464023.
PR 23-AUG-1991; US-749001.
PR 17-NOV-1993; US-154198.
PR 05-JUN-1995; US-464023.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Persica MG, Salomon DS;
DR WPI; 99-095001/08.
DR N-PSDB; V83923.
PT New antibody to human CRIPTO-related polypeptide-3 - used for the
PT detection of CRIPTO-related polypeptide-3 which is involved in the
PT regulation of the proliferation, differentiation and transformation
PT of cells
PS Example 2; Columns 17-18; 26pp; English.
CC The present sequence represents a human CRIPTO-related protein CR-1. The
CC specification describes an antibody which has binding affinity to CR-3
CC and not to CR-1. The antibodies can be used for the detection of CR-3.
CC CR-3 is a regulatory molecule involved in regulating the proliferation,
CC differentiation, and transformation of various mesenchymal and epithelial
CC cells. In addition expression of CR-3 may serve as a tumour specific
CC marker that may have applicability in the diagnosis, prognosis and
CC possible treatment of specific types of cancer.

SQ Sequence 188 AA;
Query Match 66.2%; Score 49; DB 39; Length 188;
Best Local Similarity 75.0%; Pred. No. 6.53e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 64 sqrvppmg 71
:|||||
QY 356 AQRVEPMG 363

RESULT 4
ID W29735 standard; Protein; 188 AA.
AC W29735;
DT 09-NOV-1998 (first entry)
DE Homo sapiens CRIPTO protein.
KW CRIPTO; diagnosis; prognosis; cancer; colorectal; breast.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Cleavage_site 28..29 /note= "V-A"
FT Cleavage_site 159..160 /note= "V-A"
FT Cleavage_site 111..112 /note= "V-A"
FT Cleavage_site 126..127 /note= "R-K"
FT Cleavage_site 171..174 /note= "K-K"
FT Region 79..81 /note= "R-T-T-T"
FT /note= "potential Asp glycosylation sequence"
PN US5792616-A.
PD 11-AUG-1998.
PF 29-MAY-1990; 530165.
PR 28-MAY-1990; US-530165.
PR 28-SEP-1992; US-947315.
PR 10-NOV-1994; US-337911.
PR 05-JUN-1995; US-463616.
PA (USGO) US GOVERNMENT.
PI Persico MG, Salomon DS;
DR WPI; 98-456123/39.
DR N-PSDB; V47553.
PT New antibody which binds a CRIPTO protein - is useful for screening
PT for expression of a CRIPTO protein in a tissue sample
PS Disclosure; Fig 2; 12pp; English.
CC The sequence is that of the human CRIPTO protein. Expression
CC of CRIPTO mRNA and CRIPTO protein appears to be associated with
CC various cancers, e.g. colon, breast and small cell lung carcinomas
CC and thus the sequence or antibodies raised against it may be
CC of use in the diagnosis of cancers.
SQ Sequence 188 AA;

Query Match 66.2%; Score 49; DB 34; Length 188;
Best Local Similarity 75.0%; Pred. No. 6.53e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 64 sqrvppmg 71
:|||||
QY 356 AQRVEPMG 363

RESULT 5
ID R22548 standard; Protein; 188 AA.
AC R22548;
DT 28-AUG-1992 (first entry)
DE Human CRIPTO CR-1 protein.
KW cell proliferation; tumour; CR-3; transforming growth factor;
KW epidermal growth factor.
OS Homo sapiens.
PN US7749001-A.
PD 23-FEB-1992.
PF 23-AUG-1991; 749001.
PR 23-AUG-1991; US-749001.

PA (USSH) US DEPT HEALTH & HUMAN.
PI Salomon D, Persico M;
DR WPI: 92-123675/15.
DR N-PSDB; Q22495.

PT New CRIPTO gene CR-1 and CRIPTO-related gene CR-3 genomic DNA -
PT CR-3 protein and anti-CR-3 antibodies, useful in immunoassay to
PT detect CR-3 as tumour specific marker
PS Disclosure: Page 22: 44pp; English.
CC The CR-1 genomic clone which codes for this protein can be used in
CC transgenic animals to examine the effects of overexpression of this
CC gene on development and tumorigenicity and to study regulation of
CC CR-1 gene. See also Q22494.
SQ Sequence 188 AA;

Query Match 66.2%; Score 49; DB 4; Length 188;
Best Local Similarity 75.0%; Pred. No. 6.53e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 64 sqrvppmg 71
QY 356 AQRVEPMG 363

RESULT 6

ID W32108 standard; Protein: 188 AA.
AC W32108;
DT 17-FEB-1998 (first entry)
DE Human CRIPTO protein.
KW CRIPTO gene; human; colorectal carcinoma; breast carcinoma; cancer;
KW tumour; diagnosis; prognosis; therapy; antibody.
OS Homo sapiens.

FH Key Location/Qualifiers

FT Cleavage_site 28..29

FT Modified_site 79..81

FT Cleavage_site /note= "Asn is N-glycosylated"

FT Misc_difference 126..127

FT /note= "encoded by CCT"

FT Misc_difference 87

FT /note= "encoded by GGA"

FT Misc_difference 151

FT /note= "encoded by GGC"

FT Cleavage_site 171..174

FT US5654140-A.

PD 05-AUG-1997.

PF 29-MAY-1990; 530165.

PR 29-MAY-1990; US-530165.

PR 18-SEP-1992; US-947315.

PR 10-NOV-1994; US-337911.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Persico MG, Salomon DS;

DR WPI: 97-401840/37.

DR N-PSDB; T88973.

PT New cloned CRIPTO gene - used for diagnosis and prognosis of

PT carcinoma, also antisense sequences for tumour treatment

PS Disclosure: Fig 1A-B; 13pp; English.

CC This protein comprises human CRIPTO, a protein encoded by the novel

CC CRIPTO gene that is associated with cancers such as colorectal

CC cancer and breast carcinoma. Its sequence was deduced from a

CC CRIPTO cDNA clone (see T88973) obtained from human teratocarcinoma

CC cell line NT2D1. It is a member of the epidermal growth factor

CC family. Elevated levels of CRIPTO mRNA, or of protein translated

CC from it, are diagnostic and prognostic of carcinoma, especially

CC colorectal and breast carcinoma. Recombinant CRIPTO can be

CC expressed in prokaryotic or eukaryotic host cells (see W32107), and

CC used for the identification and characterisation of specific cell

CC surface receptors, or to raise specific antibodies to assay CRIPTO

CC protein expression in immunoassays.

SQ Sequence 188 AA;

Query Match 66.2%; Score 49; DB 26; Length 188;
Best Local Similarity 75.0%; Pred. No. 6.53e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 64 sqrvppmg 71
QY 356 AQRVEPMG 363

RESULT 7

ID R32190 standard; Protein: 589 AA.
AC R32190;
DT 30-MAY-1993 (first entry)
DE Sequence encoded by the PHB synthase (phbc) gene of the
DE polyhydroxybutyrate (PHB) operon
KW Operon; polyhydroxyalkanoate; polyhydroxybutyrate synthase.
OS Alcaligenes eutrophus.

PN W09302187-A.

PD 04-FEB-1993.

PF 13-JUL-1992; U05786.

PR 19-JUL-1991; US-732243.

PA (UYMA-) UNIV MADISON JAMES.

PA (UNMS) UNIV MICHIGAN STATE.

PI Dennis DE, Foirier Y, Somerville CR;

DR WPI: 93-058785/07.

DR N-PSDB; Q36660.

PT Transgenic plants producing poly:hydroxy-alkanoate polymer(s) -

PT obtd. by transformation with DNA encoding 3-ketothiolase,

PT acetoacetyl-CoA reductase and PHA synthase

PS Disclosure: Fig 2; 70pp; English.

CC The nucleotide sequence of the PHB operon was obtained from Janes, B.

CC Hollar, J. and Dennis, D. in Dawes, E.A. (ed.) Novel Biodegradable

CC Polymers, Kluwer Academic Publishers, 175-190 (1990). It contains

CC the genes from PHB synthase, 3-ketothiolase and acetoacetyl-CoA

CC reductase. The inventors claim a transgenic plant material contg.

CC foreign DNA encoding a peptide which exhibits 3-ketothiolase activity,

CC pref. where the DNA is an open reading from between nucleotides

CC 2696-3877 (phb A gene), 842-2611 (phb C gene) or 3952-4692 (phb B

CC gene) of the Alcaligenes eutrophus PHB operon.

SQ Sequence 589 AA;

Query Match 66.2%; Score 49; DB 6; Length 589;
Best Local Similarity 40.0%; Pred. No. 6.53e+01;
Matches 4; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db 148 lsqyvdamsp 157
QY 355 VAQRVEPMGP 364

RESULT 8

ID R10681 standard; Protein: 589 AA.
AC R10681;
DT 17-APR-1991 (first entry)
DE Polyhydroxybutyrate polymerase enzyme.
KW Polyester biopolymers; polyhydroxybutyrate; polyhydroxy alkanoate;
KW beta-ketothiolase; acetoacetyl CoA reductase.
OS Alcaligenes eutrophus.

PN W09100917-A.

PD 24-JAN-1991.

PF 10-JUL-1990; U03851.

PR 10-JUL-1989; US-378155.

PA (MASI) MASSACHUSETTS INST TECH.

PI Peoples OP, Sinskey AJ;

DR WPI: 91-051341/07.

DR N-PSDB; Q10502.

PT Construction and modification of polyester bio:polymers - by

PT introduction of poly:hydroxy-butyrate and -alkanoate genes into

PT bacteria or plants

PS Disclosure: fig 4; 64pp; English.

CC This Alcaligenes eutrophus polyhydroxybutyrate (PHB) polymerase enzyme

CC is essential to the biosynthesis of PHB. The use of recombinant methods

CC for producing such enzymes, required for polyester biopolymer synthesis,

CC allows for the control and modification of the synthesis process.

CC See also Q10499-501 and

CC Q10503.

```

SQ Sequence 589 AA;
Query Match 66.2%; Score 49; DB 2; Length 589;
Best Local Similarity 40.0%; Pred. No. 6.53e+01;
Matches 4; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db 148 lsqwdamp 157
:::|::|:|
QY 355 VAQRVEPMGP 364

RESULT 9
ID R71325 standard; Protein; 672 AA.
AC R71325;
DE 21-OCT-1995 (first entry)
DT Poly-beta-hydroxyalkanoate-synthase.
KW Poly-beta-hydroxyalkanoate-synthase; transgenic plant;
KW poly-beta-hydroxyalkanoate; poly-beta-hydroxybutyrate;
KW biodegradable thermoplastic.
OS Alcaligenes eutrophus.
PN W09503472-A.
PD 23-FEB-1995.
PE 17-AUG-1994; U09265.
PR 17-AUG-1993; US-108193.
PR 06-JUN-1994; US-254357.
PA (UNMS ) UNIV MICHIGAN STATE.
PI Nawrath C, Poirier Y, Somerville CR;
DR WPI: 95-098770/13.
DR N-PSDB: R71325.
PT Transgenic plant material with plastid(s) contg. the enzymes for
PT synthesis of poly(hydroxyalkanoate(s)) - express
PT poly(hydroxybutyrate and have good growth and seed formation.
PS Claim 2; Page 62-64; 88pp; English.
CC The poly-beta-hydroxyalkanoate-synthase gene (phbC) from A.
CC eutrophus is cloned under the control of an Arabidopsis thaliana
CC seed storage protein promoter for plastid tissue-specific
CC gene expression in a transgenic plant. When expressed with the
CC 3-ketothiolase (phbA) and acetyl-CoA-reductase (phbB) genes, a
CC poly-beta-hydroxyalkanoate (PHA), specifically poly-beta-
CC hydroxybutyrate (PHB), is expressed in the transgenic plant
CC (preferably a Brassica e.g. rape). PHB and related PHAs are
CC biodegradable thermoplastics with many useful applications.
SQ Sequence 672 AA;

Query Match 66.2%; Score 49; DB 13; Length 672;
Best Local Similarity 40.0%; Pred. No. 6.53e+01;
Matches 4; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db 231 lsqwdamp 240
:::|::|:|
QY 355 VAQRVEPMGP 364

RESULT 10
ID R28754 standard; Protein; 763 AA.
AC R28754;
DT 15-APR-1993 (first entry)
DE Gastrin-binding protein.
KW Gastrin-binding protein.
KW germ: agglutinin; concanavalin A; sepharose; chromatography; receptor;
KW gastrin/cholecystokinin-C; CCK-C; acid production; cancer; gastrin;
KW colon carcinoma; ulcer; therapy; antagonist.
OS Sus scrofa domestica.
FH Key Location/Qualifiers
FT peptide 1..36
/Note= "Signal peptide"
FT protein 37..762
/Note= "Mature peptide"
FT region 37..79
/Note= "Isolated from 78 kD GBP"
FT region 133..141
/Note= "Isolated from 78 kD GBP"
FT region 236..244

Query Match 66.2%; Score 49; DB 6; Length 763;
Best Local Similarity 70.0%; Pred. No. 6.53e+01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 218 vdqlveplgp 227
| | | | | | |
QY 355 VAQRVEPMGP 364

RESULT 11
ID W41085 standard; Protein; 74 AA.
AC W41085;
DT 08-JUN-1998 (first entry)
DE Human matrix metalloproteinase hMMP-2 (aa439-512).
KW Matrix metalloproteinase; MMP-2; human; angiogenesis;
KW inhibitor; antagonist; integrin alpha-v beta-3;
KW vitronectin receptor; rheumatoid arthritis; tumour; metastasis;
KW diabetic retinopathy; macular degeneration; restenosis; therapy.
OS Homo sapiens.
PN W09745137-A1.
PD 04-DEC-1997.
PE 30-MAY-1997; U09158.
PR 31-MAY-1996; US-018733.
PR 31-MAY-1996; US-015869.
PA (SCRI ) SCRIPPS RES INST.
PI Brooks P, Cheresh DA;
DR WPI: 98-032334/03.
PT Packaging material containing polypeptide antagonist of alphav,
PT beta3 integrin - used for inhibition of angiogenesis, and for
PT treating tumours, inflammation, eye diseases etc.
PS Claim 2; Page 153; 234pp; English.

```


CC This polypeptide comprises amino acid residues 439-512 of human
CC mature matrix metalloproteinase 2 (hMMP-2). It can be produced
CC by recombinant methods such as PCR amplification of hMMP-2 coding
CC sequence and cloning into e.g. pEX-3X vector for expression in
CC E. coli as a fusion protein with glutathione-S-transferase. The
CC invention relates to the discovery that angiogenesis is mediated by
CC the specific vitronectin receptor alpha-v beta-3, and that
CC inhibition of alpha-v beta-3 function inhibits angiogenesis.
CC Claimed antagonists of alpha-v beta-3 comprise C-terminal fragments
CC (see W41083-94) of human or chicken MMP-2, fusion polypeptides,
CC cyclic or linear polypeptides (see also W41098-110), derivatised
CC polypeptides, a monoclonal antibody or organic mimetic compound.
CC The antagonists are used to inhibit angiogenesis in: inflamed
CC tissue for treatment of arthritis or rheumatoid arthritis; solid
CC tumours or metastases, particularly to induce regression or to
CC inhibit growth of tumours; and in ocular disorders such as diabetic
CC retinopathy or macular degeneration (all claimed). They can also
CC be used to treat restenosis caused by migration of smooth muscle
CC cells following angioplasty and to reduce blood supply to selected
CC tissues (Claimed). The new antagonists are highly selective for
CC angiogenesis. Only new blood vessels express alpha-v beta-3, so
CC mature vessels are unaffected, and the antagonists should be of low
CC toxicity.
SQ Sequence 74 AA;

Query Match 63.5%; Score 47; DB 29; Length 74;
Best Local Similarity 60.0%; Pred. No. 1.11e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 30 vtprdkpmgp 39
I: | ||||
QY 355 VAQRVEPMGP 364

RESULT 12
ID W41230 standard; protein; 74 AA.
AC W41230;
DT 09-JUN-1998 (first entry)
DE Alpha-v-beta-5 antagonistic peptide derived from human MMP-2.
KW Matrix metalloproteinase-2; MMP-2; alpha-v-beta-5 antagonist; treatment;
KW vitronectin receptor; inhibition; angiogenesis; tumour growth;
KW restenosis; neovascularisation.
OS Synthetic.
OS Homo sapiens.
PN W09745447-A1.
PD 04-DEC-1997.
PF 30-MAY-1997; U09099.
PR 31-MAY-1996; US-018733.
PR 31-MAY-1996; US-015869.
PA (SCRI) SCRIPPS RES INST.
PI Brooks P, Cheres DA, Friedlander M;
DR WPI; 98-041758/04.
PT Packaging material containing polypeptide antagonist of alphav,
PT beta5 integrin - used for inhibition of angiogenesis, and for
PT treating tumours, inflammation, eye diseases etc.
PS Claim 2; Page -: 117pp; English.
CC Peptides W41228-33 are derived from the mature protein of human matrix
CC metalloproteinase-2 (MMP-2) (W41226). The present peptide is derived
CC from amino acids 439-512. The peptides are able to act as alpha-v-beta-5
CC antagonists. Alpha-v-beta-5 is a vitronectin receptor. Inhibitors of
CC alpha-v-beta-5 can inhibit angiogenesis. The specification describes a
CC novel labelled package that contains an inhibitor of angiogenesis i.e. an
CC alpha-v-beta-5 antagonising polypeptide that binds to integrin
CC alpha-v-beta-5 and includes a part of the C-terminal domain of MMP-2.
CC The antagonists are used to inhibit angiogenesis in inflamed tissue, in
CC solid tumours or metastases, and in a wide range of ocular disorders
CC (e.g. diabetic or other forms of retinopathy, neovascular glaucoma, or
CC corneal transplants). They are particularly used to induce regression or
CC to inhibit growth of tumours. The alpha-v-beta-5 antagonists can also be
CC used to treat restenosis caused by migration of smooth muscle cells
CC following angioplasty and to reduce blood supply to selected tissues.
CC The antagonists particularly inhibit neovascularisation where this is
CC induced by cytokines, e.g. transforming growth factor alpha, epidermal

CC growth factor or especially vascular endothelial growth factor.
CC note: this sequence does not appear in the specification; it was created
CC using information provided.
SQ Sequence 74 AA;

Query Match 63.5%; Score 47; DB 29; Length 74;
Best Local Similarity 60.0%; Pred. No. 1.11e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 30 vtprdkpmgp 39
I: | ||||
QY 355 VAQRVEPMGP 364

RESULT 13
ID W41231 standard; protein; 108 AA.
AC W41231;
DT 09-JUN-1998 (first entry)
DE Alpha-v-beta-5 antagonistic peptide derived from human MMP-2.
KW Matrix metalloproteinase-2; MMP-2; alpha-v-beta-5 antagonist; treatment;
KW vitronectin receptor; inhibition; angiogenesis; tumour growth;
KW restenosis; neovascularisation.
OS Synthetic.
OS Homo sapiens.
PN W09745447-A1.
PD 04-DEC-1997.
PF 30-MAY-1997; U09099.
PR 31-MAY-1996; US-018733.
PR 31-MAY-1996; US-015869.
PA (SCRI) SCRIPPS RES INST.
PI Brooks P, Cheres DA, Friedlander M;
DR WPI; 98-041758/04.
PT Packaging material containing polypeptide antagonist of alphav,
PT beta5 integrin - used for inhibition of angiogenesis, and for
PT treating tumours, inflammation, eye diseases etc.
PS Claim 2; Page -: 117pp; English.
CC Peptides W41228-33 are derived from the mature protein of human matrix
CC metalloproteinase-2 (MMP-2) (W41226). The present peptide is derived
CC from amino acids 439-512. The peptides are able to act as alpha-v-beta-5
CC antagonists. Alpha-v-beta-5 is a vitronectin receptor. Inhibitors of
CC alpha-v-beta-5 can inhibit angiogenesis. The specification describes a
CC novel labelled package that contains an inhibitor of angiogenesis i.e. an
CC alpha-v-beta-5 antagonising polypeptide that binds to integrin
CC alpha-v-beta-5 and includes a part of the C-terminal domain of MMP-2.
CC The antagonists are used to inhibit angiogenesis in inflamed tissue, in
CC solid tumours or metastases, and in a wide range of ocular disorders
CC (e.g. diabetic or other forms of retinopathy, neovascular glaucoma, or
CC corneal transplants). They are particularly used to induce regression or
CC to inhibit growth of tumours. The alpha-v-beta-5 antagonists can also be
CC used to treat restenosis caused by migration of smooth muscle cells
CC following angioplasty and to reduce blood supply to selected tissues.
CC The antagonists particularly inhibit neovascularisation where this is
CC induced by cytokines, e.g. transforming growth factor alpha, epidermal
CC growth factor or especially vascular endothelial growth factor.
CC note: this sequence does not appear in the specification; it was created
CC using information provided.
SQ Sequence 108 AA;

Query Match 63.5%; Score 47; DB 29; Length 108;
Best Local Similarity 60.0%; Pred. No. 1.11e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 30 vtprdkpmgp 39
I: | ||||
QY 355 VAQRVEPMGP 364

RESULT 14
ID W41086 standard; Protein; 108 AA.
AC W41086;
DT 08-JUN-1998 (first entry)
DE Human matrix metalloproteinase hMMP-2 (aa439-546).
KW Matrix metalloproteinase; MMP-2; hMMP-2; human; angiogenesis;

KW inhibitor; antagonist; integrin alpha-v beta-3;
KW vitronectin receptor; rheumatoid arthritis; tumour; metastasis;
KW diabetic retinopathy; macular degeneration; restenosis; therapy.
OS Homo sapiens.
PN WO9745137-A1.
PD 04-DEC-1997.
PF 30-MAY-1997; U09158.
PR 31-MAY-1996; US-018733.
PR 31-MAY-1996; US-015869.
PA (SCRI) SCRIPPS RES INST.
PI Brooks P, Cheres DA;
DR WPI: 98-032334/03.
PT Packaging material containing polypeptide antagonist of alphav,
PT betag integrin - used for inhibition of angiogenesis, and for
PT treating tumours, inflammation, eye diseases etc.
PS Claim 2: Page 154: 234pp: English.
CC This polypeptide comprises amino acid residues 439-546 of human
CC mature matrix metalloproteinase 2 (hMMP-2). It can be produced
CC by recombinant methods such as PCR amplification of hMMP-2 coding
CC sequence and cloning into e.g. pGEX-3X vector for expression in
CC E. coli as a fusion protein with glutathione-S-transferase. The
CC invention relates to the discovery that angiogenesis is mediated by
CC the specific vitronectin receptor alpha-v beta-3, and that
CC inhibition of alpha-v beta-3 function inhibits angiogenesis.
CC Claimed antagonists of alpha-v beta-3 comprise C-terminal fragments
CC (see W41083-94) of human or chicken MMP-2, fusion polypeptides,
CC cyclic or linear polypeptides (see also W41098-110), derivatised
CC polypeptides, a monoclonal antibody or organic mimetic compound.
CC The antagonists are used to inhibit angiogenesis in: inflamed
CC tissue for treatment of arthritis or rheumatoid arthritis; solid
CC tumours or metastases, particularly to induce regression or to
CC inhibit growth of tumours; and in ocular disorders such as diabetic
CC retinopathy or macular degeneration (all claimed). They can also
CC be used to treat restenosis caused by migration of smooth muscle
CC cells following angioplasty and to reduce blood supply to selected
CC tissues (claimed). The new antagonists are highly selective for
CC angiogenesis. Only new blood vessels express alpha-v beta-3, so
CC mature vessels are unaffected, and the antagonists should be of low
CC toxicity.
SQ Sequence 108 AA;

Query Match 63.5%; Score 47; DB 29; Length 108;
Best Local Similarity 60.0%; Pred. No. 1.11e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 30 vtprdkmpg 39
|:|:|:|:|
QY 355 VAQRVEPMGP 364

RESULT 15
ID W04552 standard; Protein; 191 AA.
AC W04552;
DT 22-OCT-1997 (first entry)
DE Retroviral reverse transcriptase associated with Sjogren's syndrome.
KW Sjo-1; Sjogren's syndrome; biopsy; retrovirus; B; D-type; protease;
KW reverse transcriptase; JC96; human; exogenous; diagnosis;
KW autoimmune disease.
OS Human retrovirus.
PN WO9634942-A1.
PD 07-NOV-1996.
PF 03-MAY-1996; G01073.
PR 05-MAY-1995; GB-009245.
PA (BOYD/) BOYD M T.
PA (GRIF/) GRIFFITHS D J.
PA (VENA/) VENABLES P J.
PA (WEIS/) WEISS R RA.
PI Boyd MT, Griffiths DJ, Venables PJ, Weiss RA;
DR WPI: 96-506149/50.
DR N-PSDB: T38775.
PT Retroviral protease and polymerase genes - useful to develop prods.
PT for diagnosis and treatment of autoimmune diseases, e.g. Sjogren's
PT syndrome.

PS Claim 1; Fig 2; 54pp; English.
CC This reverse transcriptase is encoded by a novel human exogenous
CC retroviral cDNA clone designated JC96. JC96 was isolated using degenerate
CC primers based on Sjo-1 (see W04553). A 136 bp sequence encoding Sjo-1 was
CC isolated by RT-PCR amplification of total RNA from a Sjogren's syndrome
CC salivary gland lip biopsy. The peptide showed close homology with D and
CC B-type retroviral sequences. Clone JC96 also encodes a protease. The
CC genes and proteins are useful for diagnosis of retrovirus-associated
CC autoimmune diseases, e.g. Sjogren's syndrome.
SQ Sequence 191 AA;
Query Match 63.5%; Score 47; DB 23; Length 191;
Best Local Similarity 50.0%; Pred. No. 1.11e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 84 vncqmqpmga 93
|:|:|:|:|
QY 355 VAQRVEPMGP 364
Search completed: Thu Jul 8 18:51:03 1999
Job time : 17 secs.

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:49:48 1999; Maspar time 5.32 Seconds
Tabular output not generated. 75.254 Million cell updates/sec

Title: >US-09-041-236-2
Description: (355-364) from US09041236.ppep (21 of 45)
Perfect Score: 74
Sequence: 1 VAQRVPMGP 10

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.036; Variance 27.764; scale 0.830

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	52	70.3	546	2	F70866	1.74e+00
2	52	70.3	1571	2	T00062	hypothetical protein 1.74e+00
3	51	68.9	372	2	F70783	hypothetical protein 2.85e+00
4	50	67.6	962	2	T01688	starch debranching en 4.82e+00
5	49	66.2	166	2	T01263	hypothetical protein 7.45e+00
6	49	66.2	188	2	A30362	epidermal growth fact 7.45e+00
7	49	66.2	226	2	D35719	hisp-like nucleotide 7.45e+00
8	49	66.2	275	2	PN0511	gastrin-binding prote 7.45e+00
9	49	66.2	353	2	D71526	hypothetical protein 7.45e+00
10	49	66.2	462	2	A34341	probable hemL protein 7.45e+00
11	49	66.2	589	2	G34341	poly(3-hydroxybutyrat 7.45e+00
12	49	66.2	763	2	JC2108	long-chain-fatty-acid 7.45e+00
13	48	64.9	353	2	A54969	beta-5 GTP-binding pr 1.19e+01
14	48	64.9	763	2	A49681	long-chain-fatty-acid 1.19e+01
15	48	64.9	788	1	OQBEE3	HHLF1 protein - human 1.19e+01
16	48	64.9	846	1	OQBEC3	HQRF1 protein - human 1.19e+01
17	48	64.9	1454	2	S53398	hypothetical protein 1.19e+01
18	47	63.5	362	2	S17285	hypothetical protein 1.90e+01
19	47	63.5	450	2	A69047	NADH oxidase - Methan 1.90e+01
20	47	63.5	457	2	A64924	probable membrane pro 1.90e+01
21	47	63.5	650	1	A28153	gelatinase A (EC 3.4. 1.90e+01
22	47	63.5	675	2	S20819	collagen alpha 3(I)X 1.90e+01
23	46	62.2	127	2	S44955	lmbH protein - Strept 2.99e+01

24	46	62.2	166	2	A70391	conserved hypotheticala 2.99e+01
25	46	62.2	188	2	A39787	epidermal growth fact 2.99e+01
26	46	62.2	251	2	I39638	ribosomal protein S3 2.99e+01
27	46	62.2	376	2	S17246	chorismate synthase (2.99e+01
28	46	62.2	463	2	A64381	NADH oxidase - Methan 2.99e+01
29	46	62.2	569	2	S11035	chaperonin hsp60, tes 2.99e+01
30	46	62.2	1488	2	T02856	probable membrane pro 2.99e+01
31	45	60.8	360	2	S06280	decorin precursor - b 4.69e+01
32	45	60.8	438	2	A57667	pop-1 protein - Caeno 4.69e+01
33	45	60.8	446	2	S72889	glutamate-1-semialdeh 4.69e+01
34	45	60.8	651	2	S53415	hypothetical protein 4.69e+01
35	45	60.8	806	2	E64221	phenylalanine-tRNA 1 4.69e+01
36	45	60.8	992	2	G70950	hypothetical protein 4.69e+01
37	45	60.8	1210	2	A48001	phospholipase C (EC 3 4.69e+01
38	45	60.8	1436	2	JC5290	protein-tyrosine-phos 4.69e+01
39	45	60.8	1442	2	S72441	protein-tyrosine-phos 4.69e+01
40	45	60.8	1490	2	J73251	nonstructural polypro 4.69e+01
41	45	60.8	1612	2	JC5210	DNA (cytosine-5)-met 4.69e+01
42	45	60.8	2944	2	A54849	collagen alpha 1(VII) 4.69e+01
43	44	59.5	57	2	F35056	MHC class II histocom 7.29e+01
44	44	59.5	246	2	S38950	Ig gamma chain - mous 7.29e+01
45	44	59.5	1926	2	S01169	beta-glycosidase comp 7.29e+01

ALIGNMENTS

RESULT 1

ENTRY F70866 #type complete
TITLE probable maltase - Mycobacterium tuberculosis (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
ACCESSIONS F70866
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
#cross-references MUID:98295987
#accession F70866
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-546 #label COL
#cross-references GB:AL021246; GB:AL123456; NID:g3261507; PID:el237596;
#experimental_source strain H37Rv
GENETICS
#gene Rv2471
SUMMARY #length 546 #molecular-weight 60443 #checksum 3147
Query Match 70.3% Score 52; DB 2; Length 546;
Best Local Similarity 75.0%; Pred. No. 1.74e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 5 ORDPMPGP 12
Qy 357 QRVEPMGP 364
RESULT 2
ENTRY T00062 #type fragment
TITLE hypothetical protein KIAA0434 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man

```

DATE      22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
ACCESSIONS T00062
REFERENCE   Z14082
#authors   Ishikawa, K.; Nagase, T.; Nakajima, D.; Seki, N.; Ohira, M.;
           Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
#submission submitted to the EMBL Data Library, October 1997
#accession  T00062
#status     preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues   1-1571 #label ISH
#cross-references EMBL:AB007894; NID:d1175350; PID:d1024611
#experimental_source brain; clone HH2165
GENETICS
#note       KIAA0434
#length     1571 #checksum 5862
SUMMARY
Query Match      70.3%; Score 52; DB 2; Length 1571;
Best Local Similarity 66.7%; Pred. No. 1.74e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 896 SORLEPLGP 904
   :||:|:|:|
Qy 356 AQRVEPMGP 364

RESULT 3
ENTRY   G70783 #type complete
TITLE   hypothetical protein RV0906 - Mycobacterium tuberculosis
        (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE      17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
ACCESSIONS G70783
REFERENCE   A70500
#authors   Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
           C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
           III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
           Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
           Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
           Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
           Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
           Rajadaram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
           Skellern, S.; Squares, S.; Squires, R.; Sulston, J.E.;
           Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal   Nature (1998) 393:537-544
#title     Deciphering the biology of Mycobacterium tuberculosis from
           the complete genome sequence.
#cross-references MUID:98295987
#accession  G70783
#status     preliminary; nucleic acid sequence not shown;
           translation not shown
#molecule_type DNA
#residues   1-372 #label COL
#cross-references GB:Z73101; GB:AL123456; NID:g3261565; PID:e241999;
           PID:g1314046
#experimental_source strain H37Rv
GENETICS
#gene       RV0906
SUMMARY
#length     372 #molecular-weight 40641 #checksum 2597
Query Match      68.9%; Score 51; DB 2; Length 372;
Best Local Similarity 66.7%; Pred. No. 2.85e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 355 GQRVDPTGP 363
   :||:|:|
Qy 356 AQRVEPMGP 364

RESULT 4
ENTRY   T01688 #type complete
TITLE   starch debranching enzyme, pullulanase-type - maize

```

```

ORGANISM #formal_name Zea mays #common_name maize
DATE      19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change
ACCESSIONS T01688
REFERENCE   Z14395
#authors   Beatty, M.K.; Rahman, A.; Cao, H.; Woodman, W.; Lee, M.;
           Myers, A.M.; James, M.G.
#journal   Plant Physiol. (1999) 119:255-266
#title     Purification and molecular genetic characterization of ZPUL,
           a pullulanase-type starch-debranching enzyme from maize.
#accession  T01688
#status     preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues   1-962 #label BEA
#cross-references EMBL:AF080567; NID:g3411265; PID:g3411266
#experimental_source endosperm
GENETICS
#gene       zpul
FUNCTION
#description  hydrolyzes alpha 1,6 glucosyl linkages
SUMMARY      #length 962 #molecular-weight 106239 #checksum 5188
Query Match      67.6%; Score 50; DB 2; Length 962;
Best Local Similarity 70.0%; Pred. No. 4.62e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 48 VAQRVPEVAP 57
   ||||| |::|
Qy 355 VAQRVPEMGP 364

RESULT 5
ENTRY   T01263 #type complete
TITLE   hypothetical protein 1 - African clawed frog
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE      12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
ACCESSIONS T01263
REFERENCE   Z14286
#authors   Krol, A.; Lund, E.; Dahlberg, J.E.
#journal   EMBO J. (1985) 4:1529-1535
#title     The two embryonic UI RNA genes of xenopus laevis have both
           common and gene-specific transcription signals.
#cross-references MUID:85284944
#accession  T01263
#status     preliminary; translated from GB/EMBL/DBDJ
#molecule_type DNA
#residues   1-166 #label KRO
#cross-references EMBL:X02585; NID:g65182; PID:g65183
SUMMARY      #length 166 #molecular-weight 18233 #checksum 7801
Query Match      66.2%; Score 49; DB 2; Length 166;
Best Local Similarity 62.5%; Pred. No. 7.45e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 52 ERVQPLGP 59
   :||:|:|
Qy 357 QRVEPMGP 364

RESULT 6
ENTRY   A30362 #type complete
TITLE   epidermal growth factor-like protein CR3 - human
ALTERNATE_NAMES CRIPTO protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE      18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change
ACCESSIONS B39787; A30362
REFERENCE   A39787
#authors   Dono, R.; Montuori, N.; Rocchi, M.; De Ponti-Zilli, L.;
           Ciccocioppa, A.; Persico, M.G.
#journal   Am. J. Hum. Genet. (1991) 49:555-565
#title     Isolation and characterization of the CRIPTO autosomal gene

```

```

and its X-linked related sequence.
#cross-references MUID:91353571
#accession B39787
#status preliminary
#molecule_type DNA
#residues 1-188 #label DON
#cross-references GB:M95955; GB:M37099; NID:g339430; PID:g339431
REFERENCE A30362
#authors Ciccodicola, A.; Dono, R.; Obici, S.; Simeone, A.; Zollo, M.; Persico, M.G.
#journal EMBO J. (1989) 8:1987-1991
#title Molecular characterization of a gene of the 'EGF family' expressed in undifferentiated human NTera2 teratocarcinoma cells.
#cross-references MUID:90005403
#accession A30362
#molecule_type mRNA
#residues 1-188 #label CIC
#cross-references GB:X14253; NID:g30220; PID:g30221
KEYWORDS growth factor
SUMMARY #length 188 #molecular-weight 21168 #checksum 7175

Query Match 66.2%; Score 49; DB 2; Length 188;
Best Local Similarity 75.0%; Pred. No. 7.45e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 64 SORVPPMG 71
QY 356 AQRVEPMG 363

RESULT 7
ENTRY D35719 #type complete
TITLE hisp-like nucleotide binding protein phnL - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 05-Oct-1990 #sequence_revision 05-Oct-1990 #text_change
ACCESSIONS D35719; S56324; G65218
REFERENCE A35718
#authors Chen, C.M.; Ye, Q.Z.; Zhu, Z.; Wanner, B.L.; Walsh, C.T.
#journal J. Biol. Chem. (1990) 265:4461-4471
#title Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the phn (psb) genes involved in alkyphosphonate uptake and C-P lyase activity in Escherichia coli B.
#cross-references MUID:90170953
#accession D35719
#molecule_type DNA
#residues 1-226 #label CHE
#cross-references GB:J05260; NID:g147192; PID:g147207
REFERENCE S56314
#authors Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
#journal Nucleic Acids Res. (1995) 23:2105-2119
#title Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes.
#cross-references MUID:9334362
#accession S56324
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-226 #label BUR
#cross-references EMBL:U14003; NID:g1263172; PID:g536940
#note the nucleotide sequence was submitted to the EMBL Data Library, August 1994
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.

```

```

#cross-references MUID:97426617
#accession G65218
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-226 #label BLAT
#cross-references GB:AE000482; GB:U00096; NID:g2367349; PID:g1790534; UWGP:b4096
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene phnL
CLASSIFICATION #superfamily ATP-binding cassette homology
KEYWORDS p-loop
FEATURES
24-225 #domain ATP-binding cassette homology #label ABC\
41-48 #region nucleotide-binding motif A (P-loop)
SUMMARY #length 226 #molecular-weight 24705 #checksum 8967

Query Match 66.2%; Score 49; DB 2; Length 226;
Best Local Similarity 60.0%; Pred. No. 7.45e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 215 VADRLHPMGA 224
QY 355 VAQRVEPMGP 364

RESULT 8
ENTRY PN0511 #type fragment
TITLE gastrin-binding protein precursor - pig (fragment)
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
ACCESSIONS PN0511
REFERENCE PN0511
#authors Baldwin, G.S.; Casey, A.; Weinstock, J.
#journal Biochem. Biophys. Res. Commun. (1993) 193:560-564
#title Partial structure of the gene encoding the 78kDa gastrin binding protein encodes a close relationship with the peroxisomal trifunctional enzyme.
#cross-references MUID:93290643
#accession PN0511
#molecule_type mRNA
#residues 1-275 #label BAL
#note complete nucleotide sequence not given
GENETICS
#introns 23/1; 37/1; 60/3; 105/2; 151/3; 191/3; 226/1
CLASSIFICATION #superfamily enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA dehydrogenase homology; enoyl-CoA hydratase homology
FEATURES
1-36 #domain signal sequence #status predicted #label SIG\
37-275 #product gastrin-binding protein #status predicted
#label MAT\
62-218 #domain enoyl-CoA hydratase homology #label ECH
SUMMARY #length 275 #checksum 8422

Query Match 66.2%; Score 49; DB 2; Length 275;
Best Local Similarity 70.0%; Pred. No. 7.45e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 218 VDQLVEPLGP 227
QY 355 VAQRVEPMGP 364

RESULT 9
ENTRY D71526 #type complete
TITLE hypothetical protein Cr355 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
21-Nov-1998

```

```

ACCESSIONS D71526
REFERENCE A71570
#authors Stephens, R.S.; Kalman, S.; Lammell, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:754-759
#title Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.
#cross-references MUID:99000809
#accession D71526
#status preliminary
#molecule_type DNA
#residues 1-353 #label ARN
##cross-references GB:AE001308; GB:AE001273; NID:g3328766; PID:g3328776
##experimental_source serotype D, strain UW-3/Cx

GENETICS
#gene CT355
SUMMARY #length 353 #molecular-weight 39969 #checksum 6820
Query Match 66.2%; Score 49; DB 2; Length 353;
Best Local Similarity 87.5%; Pred. No. 7.45e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 156 VAQRVEGM 163
|||||
QY 355 VAQRVEPM 362

RESULT 10
ENTRY G70544 #type complete
TITLE probable hemL protein - Mycobacterium tuberculosis (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998

ACCESSIONS G70544
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
#cross-references MUID:98295987
#accession G70544
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-462 #label COL
##cross-references GB:Z95558; GB:AL123456; NID:g3261781; PID:e316970; PID:g2113986
##experimental_source strain H37Rv

GENETICS
#gene hemL
SUMMARY #length 462 #molecular-weight 47515 #checksum 1294
Query Match 66.2%; Score 49; DB 2; Length 462;
Best Local Similarity 60.0%; Pred. No. 7.45e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 314 VMORLAPLGP 323
|||||
QY 355 VAQRVEPMGP 364

RESULT 11

```

```

ACCESSIONS A34341 #type complete
REFERENCE poly(3-hydroxybutyrate) synthase (EC 2.3.1.-) - Alcaligenes eutrophus
#formal_name Alcaligenes eutrophus
DATE 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Mar-1998
ACCESSIONS A34341
REFERENCE A34341
#authors Peoples, O.P.; Sinskey, A.J.
#journal J. Biol. Chem. (1989) 264:15298-15303
#title Poly-beta-hydroxybutyrate (PHB) biosynthesis in Alcaligenes eutrophus H16. Identification and characterization of the PHB polymerase gene (phbC).
#cross-references MUID:89359357
#accession A34341
#status preliminary
#molecule_type DNA
#residues 1-589 #label PEO
##cross-references GB:J05003; NID:g141958; PID:g141959
##experimental_source strain H16
REFERENCE A39190
#authors Schubert, P.; Krueger, N.; Steinbuechel, A.
#journal J. Bacteriol. (1991) 173:168-175
#title Molecular analysis of the Alcaligenes eutrophus poly(3-hydroxybutyrate) biosynthetic operon: identification of the N terminus of poly(3-hydroxybutyrate) synthase and identification of the promoter.
#cross-references MUID:91100279
#accession A39190
#status preliminary
#molecule_type DNA
#residues 1-219 #label SCH
##cross-references GB:M64341; NID:g141964; PID:g141965
#note the authors translated the codon TAC for residue 120 as Thr

KEYWORDS
SUMMARY acyltransferase
#length 589 #molecular-weight 64316 #checksum 6698
Query Match 66.2%; Score 49; DB 2; Length 589;
Best Local Similarity 40.0%; Pred. No. 7.45e+00;
Matches 4; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db 148 ISQWVDAMSP 157
:::|:::|
QY 355 VAQRVEPMGP 364

RESULT 12
ENTRY JC2108 #type complete
TITLE long-chain-fatty-acid beta-oxidation multienzyme complex
ALTERNATE_NAMES alpha chain precursor, mitochondrial - human
CONTAINS 78k gastrin-binding protein
long-chain-3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.211); long-chain-enoyl-CoA hydratase (EC 4.2.1.74)
#formal_name Homo sapiens #common_name man
DATE 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 24-Sep-1998
ACCESSIONS JC2108; PC2058; S50127
REFERENCE JC2108
#authors Kamiya, T.; Aoyama, T.; Komiyama, A.; Hashimoto, T.
#journal Biochem. Biophys. Res. Commun. (1994) 199:818-825
#title Structural analysis of cDNAs for subunits of human mitochondrial fatty acid beta-oxidation trifunctional protein.
#cross-references MUID:94183263
#accession JC2108
#molecule_type mRNA
#residues 1-763 #label KAM
##cross-references GB:D16480; NID:g493657; PID:d1004457; PID:g862457
#accession PC2058
#molecule_type protein
#residues 37-64 #label KA2
#note peptide sequence from amino end of mature protein

```

```

REFERENCE      S50127
#authors      Zhang, O.X.; Baldwin, G.S.
#journal      Biochim. Biophys. Acta (1994) 1219:567-575
#title        Structures of the human cDNA and gene encoding the 78 kDa
               gastrin-binding protein and of a related pseudogene.
#cross-references MUID:95002180
#accession    S50127
#status       preliminary; translation not shown
#molecule_type mRNA
#residues     1-145, 'L', 147-151, 'L', 153-170, 'A', 172-177, 'I', 179-196,
               'VF', 199-205, 'N', 207-210, 'S', 212-575, 'P', 577-693, 'L',
               695-763 ##label ZHA
#cross-references EMBL:U04627; NID:g595366; PID:g595367
GENETICS
#gene         GDB:HADA
#cross-references GDB:434026; OMIM:600890
#map_position 2p23-2p23
COMPLEX
CLASSIFICATION heterooctamer of 4 alpha and 4 beta chains
                 superfamily enoyl-CoA hydratase/3-hydroxyacyl-CoA
                 dehydrogenase; 3-hydroxyacyl-CoA dehydrogenase homology;
                 enoyl-CoA hydratase homology
KEYWORDS      acyltransferase; carbon-oxygen lyase; fatty acid
                 beta-oxidation; heterooctamer; hydro-lyase; mitochondrial;
                 NAD: oxidoreductase
FEATURE
1-36          #domain transit peptide (mitochondrion) #status
               predicted #label TPV\
37-763        #product fatty acid beta-oxidation trifunctional
               protein, alpha chain #status experimental #label MAT\
62-218        #domain enoyl-CoA hydratase homology #label ECH\
361-640        #domain 3-hydroxyacyl-CoA dehydrogenase homology #label
               HCD\
363-391       #region beta-alpha-beta NAD nucleotide-binding fold
SUMMARY        #length 763 #molecular-weight 82959 #checksum 9088
Query Match    66.2%; Score 49; DB 2; Length 763;
Best Local Similarity 70.0%; Pred. No. 7.45e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 218 VDQVLEPLGP 227
| | | | |
QY 355 VAQRVPMGP 364

RESULT 13
ENTRY   A54969
TITLE   beta-5 GTP-binding protein - mouse
ORGANISM Mus musculus #common_name house mouse
DATE    11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change
17-Mar-1999
ACCESSIONS A54969
REFERENCE   A54969
#authors    Watson, A.J.; Katz, A.; Simon, M.I.
#journal    J. Biol. Chem. (1994) 269:22150-22156
#title      A fifth member of the mammalian G-protein beta-subunit
               family. Expression in brain and activation of the beta2
               isotype of phospholipase C.
#cross-references MUID:94350964
#accession  A54969
#status     preliminary
#molecule_type mRNA
#residues   1-353 #label WAT
#cross-references GB:L34290; NID:g1237255; PID:g557738
CLASSIFICATION superfamily GTP-binding regulatory protein beta chain; WD
               repeat homology
FEATURE
59-92        #domain WD repeat homology #label WD1\
149-182       #domain WD repeat homology #label WD3\
193-226       #domain WD repeat homology #label WD4\
279-312       #domain WD repeat homology #label WD6\
321-353       #domain WD repeat homology #status atypical #label WD7
SUMMARY        #length 353 #molecular-weight 38731 #checksum 9899

```

```

Query Match    64.9%; Score 48; DB 2; Length 353;
Best Local Similarity 66.7%; Pred. No. 1.19e+01;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 41 VABRVEALG 49
| | | | |
QY 355 VAQRVPMGP 363

RESULT 14
ENTRY   A49681
TITLE   long-chain-fatty-acid beta-oxidation multienzyme complex
CONTAINS long-chain-fatty-acid precursor, mitochondrial - rat
          long-chain-3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.211);
          long-chain-enoyl-CoA hydratase (EC 4.2.1.74)
          #formal_name Rattus norvegicus #common_name Norway rat
          07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change
          20-Mar-1998
ACCESSIONS A49681
REFERENCE   A49681
#authors    Kamiyo, T.; Aoyama, T.; Miyazaki, J.; Hashimoto, T.
#journal    J. Biol. Chem. (1993) 268:26452-26460
#title      Molecular cloning of the cDNAs for the subunits of rat
               mitochondrial fatty acid beta-oxidation multienzyme
               complex. Structural and functional relationships to other
               mitochondrial and peroxisomal beta-oxidation enzymes.
#cross-references MUID:94075334
#accession  A49681
#status     preliminary
#molecule_type mRNA; protein
#residues   1-763 #label KAM
#cross-references GB:D16478; NID:g510107; PID:d1004455; PID:g510108
#experimental_source Wistar, liver
#note       sequence extracted from NCBI backbone (NCBIN:140844,
               NCBI:P:140846)
COMPLEX      heterooctamer of 4 alpha and 4 beta chains
CLASSIFICATION superfamily enoyl-CoA hydratase/3-hydroxyacyl-CoA
                 dehydrogenase; 3-hydroxyacyl-CoA dehydrogenase homology;
                 enoyl-CoA hydratase homology
KEYWORDS      carbon-oxygen lyase; fatty acid beta-oxidation;
                 heterooctamer; hydro-lyase; mitochondrion; NAD;
                 oxidoreductase
FEATURE
62-218       #domain enoyl-CoA hydratase homology #label ECH\
361-640       #domain 3-hydroxyacyl-CoA dehydrogenase homology #label
               HCD\
363-391       #region beta-alpha-beta NAD nucleotide-binding fold
SUMMARY        #length 763 #molecular-weight 82512 #checksum 9788
Query Match    64.9%; Score 48; DB 2; Length 763;
Best Local Similarity 60.0%; Pred. No. 1.19e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 218 VDQVLEPLGP 227
| | | | |
QY 355 VAQRVPMGP 364

RESULT 15
ENTRY   QBEE3
TITLE   HHLF1 protein - human cytomegalovirus (strain AD169)
ALTERNATE_NAMES hypothetical protein TRS1
ORGANISM human cytomegalovirus, human herpesvirus 5
DATE    30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change
24-Oct-1997
ACCESSIONS C27349; S09951
REFERENCE   A92935
#authors    Weston, K.; Barrell, B.G.
#journal    J. Mol. Biol. (1986) 192:177-208
#title      Sequence of the short unique region, short repeats, and part
               of the long repeats of human cytomegalovirus.
#cross-references MUID:87169717
#accession  C27349

```

```
##molecule_type DNA
##residues 1-788 ##label WES
##cross-references EMBL:X04650
##experimental_source strain AD169
REFERENCE S09749
#authors Chee, M.S.; Bankier, A.T.; Beck, S.; Bohnl, R.; Brown, C.M.;
Cerny, R.; Horsnell, T.; Hutchinson III, C.A.; Kouzarides,
T.; Martignetti, J.A.; Preddie, E.; Satchwell, S.C.;
Tomlinson, P.; Weston, K.M.; Bartell, B.G.
#journal Curr. Top. Microbiol. Immunol. (1990) 154:125-169
#title Analysis of the protein-coding content of the sequence of
human cytomegalovirus strain AD169.
#cross-references MUID:90269039
#accession S09951
#status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-788 ##label CHE
##cross-references EMBL:X17403; NID:g59591; PID:gl780968
##experimental_source strain AD169
##note this sequence was submitted to the EMBL Data Library,
December 1989
GENETICS
#gcnc HHLF1
CLASSIFICATION #superfamily cytomegalovirus HQRF1 protein
SUMMARY #length 788 #molecular-weight 83981 #checksum 7858
Query Match 64.9%; Score 48; DB 1; Length 788;
Best Local Similarity 75.0%; Pred. No. 1.19e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 60 HRVEPRGP 67
QY 357 QRVEPMGP 364
:|||||
```

Search completed: Thu Jul 8 18:50:00 1999
Job time : 12 secs.

WQRLH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:48:46 1999; MasPar time 3.34 Seconds
Tabular output not generated. 84.678 Million cell updates/sec

Title: >US-09-041-236-2
Description: (355-364) from US09041236.pep (21 of 45)
Perfect Score: 74
Sequence: 1 VQORVPMGP 10

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 23.705; Variance 24.651; scale 0.962

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description	Pred. No.
1	51	68.9	1 Y234_MYCTU	HYPOTHETICAL 40.6 KD P	8.11e+01
2	49	66.2	1 CR11_HUMAN	TERATOCARCINOMA-DERIVE	2.39e+00
3	49	66.2	1 PHNL_ECOLI	PHOSPHONATES TRANSPORT	2.39e+00
4	49	66.2	1 GSA_MYCTU	GLUTAMATE-1-SEMIALDEHY	2.39e+00
5	49	66.2	1 PHRC_ALCEU	POLY-BETA-HYDROXYBUTYR	2.39e+00
6	49	66.2	1 ECHA_HUMAN	MITOCHONDRIAL TRIFUNCT	2.39e+00
7	49	66.2	1 YDOB_SCHPO	HYPOTHETICAL 234.0 KD	2.39e+00
8	48	64.9	1 GBB5_HUMAN	GUANINE NUCLEOTIDE-BIN	4.05e+00
9	48	64.9	1 GBB5_MOUSE	GUANINE NUCLEOTIDE-BIN	4.05e+00
10	48	64.9	1 ECHA_RAT	MITOCHONDRIAL TRIFUNCT	4.05e+00
11	48	64.9	1 TRS1_HCMVA	HYPOTHETICAL PROTEIN H	4.05e+00
12	48	64.9	1 IRS1_HCMVA	HYPOTHETICAL PROTEIN H	4.05e+00
13	47	63.5	1 OGRE_DROME	OGRE LOCUS PROTEIN.	6.82e+00
14	47	63.5	1 YDHE_ECOLI	HYPOTHETICAL 49.4 KD P	6.82e+00
15	47	63.5	1 COG2_HUMAN	72 KD TYPE IV COLLAGEN	6.82e+00
16	47	63.5	1 CA39_CHICK	COLLAGEN ALPHA 3(IX) C	6.82e+00
17	47	63.5	1 TOP2_CAEEL	PROBABLE DNA TOPOISOME	6.82e+00
18	46	62.2	1 R1M1_AQUAE	TERATOCARCINOMA-DERIVE	1.14e+01
19	46	62.2	1 CR12_HUMAN	PROBABLE 16S RNA PROC	1.14e+01
20	46	62.2	1 RS3_ANAAB	30S RIBOSOMAL PROTEIN	1.14e+01
21	46	62.2	1 Y176_HUMAN	HYPOTHETICAL PROTEIN K	1.14e+01
22	46	62.2	1 AROC_YEAST	CHORISMATE SYNTHASE (E	1.14e+01
23	46	62.2	1 NAOX_METJA	PUTATIVE NADH OXIDASE	1.14e+01

24	46	62.2	569	1	CH63_HELVI	63 KD CHAPERONIN PRECU	1.14e+01
25	45	60.8	360	1	PGS2_BOVIN	BONE PROTEOGLYCAN II P	1.88e+01
26	45	60.8	438	1	POP1_CAEEL	POP-1 PROTEIN.	1.88e+01
27	45	60.8	446	1	GSA_MYCTU	GLUTAMATE-1-SEMIALDEHY	1.88e+01
28	45	60.8	651	1	CORO_YEAST	CORONIN-LIKE PROTEIN.	1.88e+01
29	45	60.8	806	1	SVFB_MYCGE	PHENYLALANYL-TRNA SYNT	1.88e+01
30	45	60.8	819	1	Y083_NPVOP	HYPOTHETICAL 91.1 KD P	1.88e+01
31	45	60.8	943	1	SYL_PSEFL	ISOLEUCYL-TRNA SYNTHET	1.88e+01
32	45	60.8	962	1	UL52_PVRKA	DNA HELICASE/PRIMASE C	1.88e+01
33	45	60.8	1374	1	YOS6_CAEEL	HYPOTHETICAL SUBTILASE	1.88e+01
34	45	60.8	1430	1	PCP2_HUMAN	PROTEIN-TYROSINE PHOSP	1.88e+01
35	45	60.8	2944	1	CA17_HUMAN	COLLAGEN ALPHA 1(VII)	1.88e+01
36	44	59.5	212	1	MAD2_PETHY	FLORAL HOMEOTIC PROTEI	3.07e+01
37	44	59.5	222	1	RS2_METJA	30S RIBOSOMAL PROTEIN	3.07e+01
38	44	59.5	266	1	APAI_RABIT	APOLIPROTEIN A-I PRE	3.07e+01
39	44	59.5	302	1	SPEE_HUMAN	SPERMIDINE SYNTHASE (E	3.07e+01
40	44	59.5	330	1	GCAA_MOUSE	IG GAMMA-2A CHAIN C RE	3.07e+01
41	44	59.5	384	1	RN_DROME	GTASE ACTIVATING PROT	3.07e+01
42	44	59.5	399	1	GCAM_MOUSE	IG GAMMA-2A CHAIN C RE	3.07e+01
43	44	59.5	874	1	POLL_HUMAN	RETROVIRUS-RELATED POL	3.07e+01
44	44	59.5	1426	1	RHSD_ECOLI	RHSD PROTEIN PRECURSOR	3.07e+01
45	44	59.5	1926	1	LPH_RABIT	LACTASE-PHLORIZIN HYDR	3.07e+01

ALIGNMENTS

RESULT	ID	Y234_MYCTU	STANDARD:	PRT:	372 AA.
AC	Q10562:				
DT	01-OCT-1996	(REL. 34, CREATED)			
DT	01-OCT-1996	(REL. 34, LAST SEQUENCE UPDATE)			
DT	15-DEC-1998	(REL. 37, LAST ANNOTATION UPDATE)			
DE	HYPOTHETICAL 40.6 KD PROTEIN CY31.34 PRECURSOR.				
GN	MTCY31.34.				
OS	MYCOBACTERIUM TUBERCULOSIS.				
OC	BACTERIA: FIRMICUTES: ACTINOBACTERIA: ACTINOBACTERIDAE:				
OC	ACTINOMYCETALES: CORYNEBACTERIINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.				
NC	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-H37RV:				
RX	MEDLINE: 98295987.				
RA	COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D., GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAFA F., BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R., DAVIES R., FELTWEIL T., FELTWEIL T., GENTILES S., HAMLIN N., HOLROYD S., HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L., OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J., RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES J.E., TAYLOR K., WHITEHEAD S., BARRELL B.G.:				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."				
RL	NATURE 393:537-544(1998).				
CC	-1- SIMILARITY: TO K.PNEUMONIAE ROMA.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL: Z73101; E241999; -				
DR	HYPOTHETICAL PROTEIN; SIGNAL.				
FT	SIGNAL 1 33				
FT	CHAIN 34 372				
SO	SEQUENCE 372 AA; 40641 MW; 6A5EFD69 CRC32;				

Query Match 68.9%; Score 51; DB 1; Length 372;
Best Local Similarity 66.7%; Pred No. 8.11e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 355 GORVDPTGP 363

OY 356 AORVEPMGP 364
:|||||

RESULT 2
ID CRIL_HUMAN STANDARD; PRT; 188 AA.
AC P13385;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE TERATOCARCINOMA-DERIVED GROWTH FACTOR 1 (EPIDERMAL GROWTH FACTOR-LIKE
DE CRYPTO PROTEIN CRI) (CRYPTO-1 GROWTH FACTOR) (CRGF).
GN TDGF1 OR CRYPTO.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90005403.
RA CICCOCICOLA A., DONO R., OBICI S., ZOLLO M., PERSICO M.G.;
RT "Molecular characterization of a gene of the 'EGF family' expressed
in undifferentiated human NTera2 teratocarcinoma cells.";
RL EMO J. 8:1987-1991(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91353571.
RA DONO R., MONTUORI N., ROCCHI M., DE PONTI-ZILLI L., CICCOCICOLA A.,
RA PERSICO M.G.;
RT "Isolation and characterization of the CRYPTO autosomal gene and its
X-linked related sequence.";
RL AM. J. HUM. GENET. 49:555-565(1991).
CC -!- FUNCTION: COULD PLAY A ROLE IN THE DETERMINATION OF THE EPIBLASTIC
CELLS THAT SUBSEQUENTLY GIVE RISE TO THE MESODERM.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN GASTRIC AND
COLORECTAL CARCINOMAS THAN IN THEIR NORMAL COUNTERPARTS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M65955; G339431; -.
DR EMBL; X14253; G30221; -.
DR PIR; A30362; A30362.
DR PIR; B39787; B39787.
DR MIW; 187395; -.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR HSPF; P00740; IIXA.
KW GROWTH FACTOR; EGF-LIKE DOMAIN; GLYCOPROTEIN.
FT DOMAIN 78 107
FT DISULFID 82 89 EGF-LIKE.
FT DISULFID 83 95 BY SIMILARITY.
FT DISULFID 97 106 BY SIMILARITY.
FT CARBOHYD 79 79 POTENTIAL.
SQ SEQUENCE 188 AA; 21168 MW; EBC57F6A CRC32;

Query Match 66.2%; Score 49; DB 1; Length 188;
Best Local Similarity 75.0%; Pred. No. 2.39e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 64 SORVPPMG 71
:|||||

OY 356 AORVEPMGP 363

RESULT 3
ID PHNL_ECOLI STANDARD; PRT; 226 AA.
AC P16679;

DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PHOSPHONATES TRANSPORT ATP-BINDING PROTEIN PHNL.
GN PHNL.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12.
RC STRAIN=K12; 91193228.
RA MAKINO K., KIM S.K., SHINAGAWA H., AMEMURA M., NAKATA A.;
RT "Molecular analysis of the cryptic and functional phn operons for
phosphate use in Escherichia coli K-12";
RL J. BACTERIOL. 173:2665-2672(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RC MEDLINE; 95334362.
RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
RA BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
region from 92.8 through 100 minutes.";
RL NUCLEIC ACIDS RES. 23:2105-2119(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=B;
RC MEDLINE; 90170953.
RA CHEN C.-M., YE Q.-Z., ZHU Z., WANNER B.L., WALSH C.T.;
RT "Molecular biology of carbon-phosphorus bond cleavage. Cloning and
sequencing of the phn (psid) genes involved in alkylphosphonate
uptake and C-P lyase activity in Escherichia coli B.";
RL J. BIOL. CHEM. 265:4461-4471(1990)
CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
FOR ALKYLPHOSPHONATES; PROBABLY RESPONSIBLE FOR ENERGY COUPLING
TO THE TRANSPORT SYSTEM.
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
(ABC TRANSPORTERS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90227; G216601; -.
DR EMBL; U14003; G536940; -.
DR EMBL; AE000482; G1790534; -.
DR EMBL; J05260; G147207; -.
DR PIR; D35719; D35719.
DR ECGENE; EG10721; PHNL.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PFAM; PF00005; ABC_tran; 1.
KW ALKYLPHOSPHONATE UPTAKE; TRANSPORT; ATP-BINDING.
FT NE_BIND 41 48 ATP (BY SIMILARITY).
SQ SEQUENCE 226 AA; 24705 MW; 42E3F6E5 CRC32;

Query Match 66.2%; Score 49; DB 1; Length 226;
Best Local Similarity 60.0%; Pred. No. 2.39e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 215 VADRLHPMGA 224
:|:|:|:

OY 355 VAQRVEPMGP 364

RESULT 4
ID GSA_MYCTU STANDARD; PRT; 462 AA.
AC O06390;
DT 15-DEC-1998 (REL. 37, CREATED)

```

DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUTAMATE-1-SEMIALDEHYDE 2,1-AMINOTRANSFERASE (EC 5.4.3.8) (GSA)
DE (GLUTAMATE-1-SEMIALDEHYDE AMINOTRANSFERASE) (GSA-AT).
GN HML OR RV0524 OR MTCY25D10.03.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; MYCOBACTERIACEAE; MYCOBACTERIUM.
OC ACTINOMYCETALES; CORYNEBACTERIINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 98295987.
RA COLE S.T., BROSH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA BACOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWELL T., GENTILES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SOARES S., SOARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL NATURE 393:537-544(1998).
CC -!- CATALYTIC ACTIVITY: (S)-4-AMINO-5-OXOPENTANOATE =
CC 5-AMINOLEVULINATE.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J05003; G141959; -
DR EMBL; M64341; G141965; -
DR PIR; A34341; A34341.
DR PIR; PF00561; abhydrolase; 1.
KW PHB BIOSYNTHESIS; TRANSFERASE; ACYLTRANSFERASE.
FT ACT_SITE 319 POTENTIAL.
SQ SEQUENCE 589 AA; 64316 MW; FCDC5F11 CRC32;
Query Match 66.2%; Score 49; DB 1; Length 589;
Best Local Similarity 40.0%; Pred. No. 2.39e+00;
Matches 4; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
Db 148 ISOWDAMSP 157
Qy 355 VAQVPEMGP 364
::|::|:|:|
RESULT 6
ID ECHA_HUMAN STANDARD; PRT: 763 AA.
AC P40939;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA SUBUNIT PRECURSOR (TP-ALPHA)
DE [CONTAINS: LONG-CHAIN ENOYL-COA HYDRATASE (EC 4.2.1.17); LONG CHAIN
DE 3-HYDROXYACYL-COA DEHYDROGENASE (EC 1.1.1.35)].
GN HADHA OR HADH.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94183263.
RA KAMIJO T., AOYAMA T., MIYAZAKI J.I., HASHIMOTO T.;
RT "Structural analysis of cDNAs for subunits of human mitochondrial
RT fatty acid beta-oxidation trifunctional protein."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 199:818-825(1994).
RN [2]
RP VARIANT AFLP GLU-510.
RX MEDLINE; 95148633.
RA SIMS H.F., BRACKETT J.C., POWELL C.K., TREEM W.R., HALE D.E.,
RA BENNETT M.J., GIBSON B., SHAPIRO S., STRAUSS A.W.;
RT "The molecular basis of pediatric long chain 3-hydroxyacyl-CoA
RT dehydrogenase deficiency associated with maternal acute fatty liver
RT of pregnancy."

```

```

RL PROC. NATL. ACAD. SCI. U.S.A. 92:841-845(1995).
CC -!- CATALYTIC ACTIVITY: (3S)-3-HYDROXYACYL-COA - TRANS-2(OR 3)-ENOYL
CC -COA + H(2)O.
CC -!- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-COA + NAD(+) -> 3-OXOACYL-COA +
CC NADH.
CC -!- PATHWAY: CATALYZES TWO STEPS OF THE FATTY ACID BETA-OXIDATION
CC CYCLE.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- DISEASE: DEFECTS IN HADHA ARE A CAUSE OF MATERNAL ILLNESS OCCURRING
CC DURING PREGNANCIES WITH AFFECTED FETUSES. THIS DISEASE IS
CC CHARACTERIZED BY SUDDEN UNEXPLAINED INFANT DEATH OR HYPOGLYCEMIA
CC AND ABNORMAL LIVER ENZYMES (REYE-LIKE SYNDROME).
CC -!- SIMILARITY: IN THE N-TERMINAL, TO ENOYL-COA HYDRATASE.
CC -!- SIMILARITY: IN THE CENTRAL SECTION, TO 3-HYDROXYACYL-COA
CC DEHYDROGENASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D16480; G862457; -.
DR PIR: JC2108; JC2108.
DR MIN: G00890; -.
DR PROSITE: PS00067; 3HCDH; 1.
DR PROSITE: PS00166; ENOYL_COA_HYDRATASE; 1.
DR PFAM: PF00378; ECH; 1.
DR PFAM: PF00725; 3HCDH; 1.
DR HSP: P14604; 2DUB.
KW FATTY ACID METABOLISM; MULTIFUNCTIONAL ENZYME; OXIDOREDUCTASE; NAD;
KW LYASE; MITOCHONDRION; TRANSIT PEPTIDE; DISEASE MUTATION.
FT TRANSIT 1 36 MITOCHONDRION (POTENTIAL).
FT CHAIN 37 763 MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA
FT ACT_SITE 151 151 ACTIVATES A WATER MOLECULE (BY
FT ACT_SITE 173 173 SIMILARITY).
FT VARIANT 510 510 PROVIDES THE ALPHA-PROTON (BY
FT SEQUENCE 763 AA; 82959 MW; CE72046C CRC32;
Query Match 66.2%; Score 49; DB 1; Length 763;
Best Local Similarity 70.0%; Pred. No. 2.39e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 218 VDQLVEPLGP 227
QY 355 VAQRVPMGP 364
| | | | |
| | | | |

RESULT 7
ID YDOB_SCHPO STANDARD; PRT; 2052 AA.
AC O13731;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 234.0 KD PROTEIN C15A10.11 IN CHROMOSOME 1.
GN SPAC15A10.11.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETACEAE; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:972;
RA MURPHY L., HARRIS D., WOOD V., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: SOME, TO YEAST UBR1.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z97208; E325335; -.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 1024 1044 POTENTIAL.
SQ SEQUENCE 2052 AA; 234041 MW; 18FA5388 CRC32;
Query Match 66.2%; Score 49; DB 1; Length 2052;
Best Local Similarity 66.7%; Pred. No. 2.39e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 278 VAQRVDSIG 286
QY 355 VAQRVPMGP 363
| | | | |
| | | | |

RESULT 8
ID GBB5_HUMAN STANDARD; PRT; 353 AA.
AC O14775;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT 5 (TRANSDUCIN BETA
DE CHAIN 5).
GN GNB5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA JONES P.G., LOMBARDI S.J., COCKETT M.I.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE
CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE
CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-
CC EFFECTOR INTERACTION.
CC -!- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF017656; G2570404; -.
DR PROSITE: PS00678; WD_REPEATS; 3.
DR PFAM: PF00400; G-beta; 6.
DR HSP: P04901; IGP2.
KW TRANSDUCER; REPEAT; WD REPEAT; MULTIGENE FAMILY.
FT REPEAT 61 91 WD1.
FT REPEAT 103 133 WD2.
FT REPEAT 151 181 WD3.
FT REPEAT 193 225 WD4.
FT REPEAT 237 267 WD5.
FT REPEAT 281 311 WD6.
FT REPEAT 323 353 WD7.
SQ SEQUENCE 353 AA; 38759 MW; 7A27CA98 CRC32;
Query Match 64.9%; Score 48; DB 1; Length 353;
Best Local Similarity 66.7%; Pred. No. 4.05e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

Db 41 VAERVEALG 49
|:|:|:|:|
QY 355 VAORVEPMG 363

RESULT 9
ID GBB5_MOUSE STANDARD; PRT; 353 AA.
AC P54314; Q35354;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT 5 (TRANSDUCIN BETA CHAIN 5).
GN GNB5.
OS MUS MUSCULUS (MUSCLE), AND RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RN SEQUENCE FROM N.A.
RC SPECIES-MOUSE; TISSUE-BRAIN;
RX MEDLINE; 94350964.
RA WATSON A.J., KATZ A., SIMON M.I.;
RT "A fifth member of the mammalian G-protein beta-subunit family. Expression in brain and activation of the beta 2 isotype of phospholipase C.";
RT J. BIOL. CHEM. 269:22150-22156(1994).
RN [2]
RN SEQUENCE FROM N.A.
RP SPECIES-RAT; STRAIN=SPRAGUE-DAWLEY; TISSUE-BRAIN;
RA LEVAY K., SLEPAK V.Z.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RN SEQUENCE OF 11-198 FROM N.A.
RC SPECIES-RAT; STRAIN=SPRAGUE-DAWLEY;
RA BETTY M., HARNISH S.W., RHODES K.J., COCKETT M.I.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-EFFECTOR INTERACTION.
CC -!- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC -!- TISSUE SPECIFICITY: BRAIN AND AT MUCH REDUCED LEVELS IN KIDNEY.
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC EMBL; L34290; G557738; -
DR EMBL; AF001953; G2182130; -
DR EMBL; AF022086; G3582396; -
DR MGD; MGI:101848; GNB5.
DR PROSITE; PS00678; WD-REPEATS; 3.
DR PFAM; PF00400; G-beta; 6.
DR HSSP; P04901; 1GP2.
KW TRANSDUCER; REPEAT; WD REPEAT; MULTIGENE FAMILY.
FT REPEAT 61 WD1.
FT REPEAT 103 WD2.
FT REPEAT 151 WD3.
FT REPEAT 181 WD3.
FT REPEAT 193 WD4.
FT REPEAT 237 WD5.
FT REPEAT 281 WD6.
FT REPEAT 323 WD6.
SQ SEQUENCE 353 AA; 38731 MW; 6BD0144F CRC32;

Query Match 64.9%; Score 48; DB 1; Length 353;
Best Local Similarity 66.78; Pred. No. 4.05e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 41 VAERVEALG 49
|:|:|:|:|
QY 355 VAORVEPMG 363

RESULT 10
ID ECHA_RAT STANDARD; PRT; 763 AA.
AC Q64428;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA SUBUNIT PRECURSOR (TP-ALPHA) [CONTAINS: LONG-CHAIN ENOYL-COA HYDRATASE (EC 4.2.1.17); LONG CHAIN 3-HYDROXYACYL-COA DEHYDROGENASE (EC 1.1.1.35)].
GN HADHA.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-WISTAR;
RX MEDLINE; 94075334.
RA KAMIJO T., AOYAMA T., MIYAZAKI J., HASHIMOTO T.;
RT "Molecular cloning of the cDNAs for the subunits of rat mitochondrial fatty acid beta-oxidation multienzyme complex. Structural and functional relationships to other mitochondrial and peroxisomal beta-oxidation enzymes.";
RT J. BIOL. CHEM. 268:26452-26460(1993).
RL -!- CATALYTIC ACTIVITY: (3S)-3-HYDROXYACYL-COA = TRANS-2(OR 3)-ENOYL -COA + H(2)O
CC -!- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-COA + NAD(+) = 3-OXOACYL-COA + NADH.
CC -!- PATHWAY: CATALYZES TWO STEPS OF THE FATTY ACID BETA-OXIDATION CYCLE.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- SIMILARITY: IN THE N-TERMINAL, TO ENOYL-COA HYDRATASE.
CC -!- SIMILARITY: IN THE CENTRAL SECTION, TO 3-HYDROXYACYL-COA DEHYDROGENASE.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC EMBL; D16478; G510108; -
DR PROSITE; PS00067; 3HCDH; 1.
DR PROSITE; PS00166; ENOYL-COA-HYDRATASE; 1.
DR PFAM; PF00378; ECH; 1.
DR HSSP; PF00725; 3HCDH; 1.
DR HSSP; P14604; 2DUB.
KW FATTY ACID METABOLISM; MULTIFUNCTIONAL ENZYME; OXIDOREDUCTASE; NAD; LYASE; MITOCHONDRIUM; TRANSIT PEPTIDE.
FT TRANSIT 1 36
FT CHAIN 37 763
FT MITOCHONDRIUM (POTENTIAL).
FT SUBUNIT
FT ACT_SITE 151 151
FT ACT_SITE 173 173
FT ACT_SITE 173 173
FT PROVIDES THE ALPHA-PROTON (BY SIMILARITY).
SQ SEQUENCE 763 AA; 82512 MW; E4367124 CRC32;

Query Match 64.9%; Score 48; DB 1; Length 763;
Best Local Similarity 60.0%; Pred. No. 4.05e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 218 VQOLVDPGLP 227
|:|:|:|:|
QY 355 VAORVEPMG 364

```

RA WESTON K., BARRELL B.G.;
RT "Sequence of the short unique region, short repeats, and part of the
RL long repeats of human cytomegalovirus.";
RN J. MOL. BIOL. 192:177-208(1986).
[2]
RP COMPLETE GENOME.
RX MEDLINE; 90269039.
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTIGNETTI J.A.,
RA PREDDIE E., SATCHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RN cytomegalovirus strain AD169.";
RL CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC -!- SIMILARITY: BELONGS TO US22 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17403; G59759; -
DR EMBL; X04650; G59803; -
DR PIR; C26078; QQBEC3.
DR PIR; S09914; S09914.
DR HSP; P41249; IAKP.
KW HYPOTHETICAL PROTEIN.
FT CARBOHYD 76 76
FT CARBOHYD 118 118
FT CARBOHYD 223 223
SQ SEQUENCE 846 AA; 91048 MW; 2B95EAE3 CRC32;

Query Match 64.9%; Score 48; DB 1; Length 846;
Best Local Similarity 75.0%; Pred. No. 4.05e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 60 HRVEPRGP 67
Qy 357 QRVEPMGP 364
:|||||

RESULT 13
ID OGFE_DROME STANDARD; PRT; 362 AA.
AC P27716;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE OGFE LOCUS PROTEIN.
GN OGFE OR L1(OGFE.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE; 91169249.
RA WATANABE T., KANKEL D.R.;
RT "Molecular cloning and analysis of l(1)ogre, a locus of Drosophila
RT melanogaster with prominent effects on the postembryonic development
RT of the central nervous system.";
RL GENETICS 126:1033-1044(1990).
CC -!- FUNCTION: PROMINENT EFFECTS ON THE POSTEMBRYONIC DEVELOPMENT
CC OF THE CENTRAL NERVOUS SYSTEM.
CC -!- TISSUE SPECIFICITY: ALSO DETECTED IN TISSUES OUTSIDE OF THE CNS.
CC -!- SIMILARITY: TO DROSOPHILA PASSOVER AND C.ELEGANS UNC-7.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

RA WESTON K., BARRELL B.G.;
RT "Sequence of the short unique region, short repeats, and part of the
RL long repeats of human cytomegalovirus.";
RN J. MOL. BIOL. 192:177-208(1986).
[2]
RP COMPLETE GENOME.
RX MEDLINE; 90269039.
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTIGNETTI J.A.,
RA PREDDIE E., SATCHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RN cytomegalovirus strain AD169.";
RL CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC -!- SIMILARITY: BELONGS TO US22 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17403; G59759; -
DR EMBL; X04650; E3311; ALT_SEQ.
DR PIR; C27349; QQBEE3.
DR PIR; S09951; S09951.
DR HSP; P41249; IAKP.
KW HYPOTHETICAL PROTEIN.
FT CARBOHYD 76 76
FT CARBOHYD 118 118
FT CARBOHYD 223 223
SQ SEQUENCE 788 AA; 83981 MW; A05DA652 CRC32;

Query Match 64.9%; Score 48; DB 1; Length 788;
Best Local Similarity 75.0%; Pred. No. 4.05e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 60 HRVEPRGP 67
Qy 357 QRVEPMGP 364
:|||||

RESULT 12
ID IRS1_HCMVA STANDARD; PRT; 846 AA.
AC P09715;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN HORF1.
GN IRS1.
OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87169717.

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X61180; G7481; -
DR PIR; S17285; S17285;
DR FLYBASE; FBgn0004646; ogre.
DR PFAM; PF00876; Ogre; 1
KW TRANSMEMBRANE; GLYCOPROTEIN.
FT DOMAIN 1 287
FT TRANSMEM 288 308
FT TRANSMEM 309 362
FT CARBOHYD 146 146
SQ SEQUENCE 362 AA; 42582 MW; E7783E69 CRC32;

Query Match 63.5%; Score 47; DB 1; Length 362;
Best Local Similarity 83.3%; Pred. No. 6.82e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 223 ORVDPM 228
Qy 357 QRVPEM 362

RESULT 14
ID YDHE-ECOLI STANDARD; PRT; 457 AA.
AC P27340; E77765; P77276;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 49.4 KD PROTEIN IN RIBC-PYKE INTERGENIC REGION.
GN YDHE.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.
RT "The complete genome sequence of Escherichia coli K-12."
RL SCIENCE 277:1453-1474(1997).

[2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA HENSEL M., SHEA J.E., BAUMLER A., GLEESON C., BLATTNER F.R.,
RA HOLDEN D.W.
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA AJBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORTUCHI T.,
RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K., KASAI H.,
RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,
RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,
RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,
RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.,
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE OF 1-230 FROM N.A.
RC STRAIN-K12 / RR28;
RA EBERHARDT S.M.R., RICHTER G., GIMBEL W., WERNER T., BACHER A.;
RL SUBMITTED (NOV-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP IDENTIFICATION.
RA RUDD K.E.;
RL UNPUBLISHED OBSERVATIONS (JUN-1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE UPF0013 FAMILY. STRONG, TO H.INFLUENZAE
CC H11612.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AE000261; G1787953; -
DR EMBL; U68703; G1549276; -
DR EMBL; D90809; G1742737; -
DR EMBL; D90810; G1742747; -
DR EMBL; X69109; -; NOT_ANNOTATED_CDS.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 11 31
FT TRANSMEM 53 73
FT TRANSMEM 93 113
FT TRANSMEM 127 147
FT TRANSMEM 161 181
FT TRANSMEM 189 209
FT TRANSMEM 243 263
FT TRANSMEM 277 297
FT TRANSMEM 314 334
FT TRANSMEM 350 370
FT TRANSMEM 387 407
FT TRANSMEM 418 438
FT TRANSMEM 438 457
FT CONFLICT 32 32 D -> S (IN REF. 1 AND 2).
FT CONFLICT 100 100 V -> A (IN REF. 4).
SQ SEQUENCE 457 AA; 49447 MW; 32507CBB CRC32;

Query Match 63.5%; Score 47; DB 1; Length 457;
Best Local Similarity 100.0%; Pred. No. 6.82e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 413 VEPMPG 418
Qy 359 VEPMPG 364

RESULT 15
ID COG2-HUMAN STANDARD; PRT; 560 AA.
AC P08253;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE 72 KD TYPE IV COLLAGENASE PRECURSOR (EC 3.4.24.24) (72 KD GELATINASE)
DE (MATRIX METALLOPROTEINASE-2) (MMP-2) (GELATINASE A) (TBE-1).
GN MMP2 OR CLG4A.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE OF 19-660 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 88198218.
RA COLLIER I.E., WILHELM S.M., EISEN A.2., WARMER B.L., GRANT G.A.,
RA SELTZER J.L., KRONBERGER A., HE C., BAUER E.A., GOLDBERG G.I.;
RT "H-ras oncogene-transformed human bronchial epithelial cells (TBE-1)
RT secrete a single metalloprotease capable of degrading basement
RT membrane collagen.";
RL J. BIOL. CHEM. 263:6579-6587(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91236162.
RA COLLIER I.E., BRUNS G.A.P., GOLDBERG G.I., GERHARD D.S.;
RT "On the structure and chromosome location of the 72- and 92-kDa human
RT type IV collagenase genes.";
RL GENOMICS 9:429-434(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90293047.
RA HUHTALA P., CHOW L.T., TRYGGVASON K.;

"Structure of the human type IV collagenase gene.";
J. BIOL. CHEM. 265:11077-11082(1990).
(4)
RN SEQUENCE OF 1-51 FROM N.A.
RP MEDLINE: 90228972
RX HUHTALA P., EDDY R.L., FAN Y.S., BYERS M.G., SHOWS T.B.,
RA TRYGGVASON K.;
RT Completion of the primary structure of the human type IV collagenase
RT preproenzyme and assignment of the gene (CLG4) to the q21 region of
RT chromosome 16.";
RL GENOMICS 6:554-559(1990).
RN [5]
RN X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 443-660.
RP MEDLINE: 96069777.
RX LIBSON A.M., GITTIS A.G., COLLIER I.E., MARMER B.L., GOLDBERG G.I.,
RA LATTMAN E.E.;
RT "Crystal structure of the haemopexin-like C-terminal domain of
RT gelatinase A";
RL NAT. STRUCT. BIOL. 2:938-942(1995).
RN [6]
RN X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 458-660.
RX MEDLINE: 96140723
RA GOHLKE U., GOMIS-RUTH F.X., CRABBE T., MURPHY G., DOCHERTY A.J.,
RA BODE W.;
RT "The C-terminal (haemopexin-like) domain structure of human
RT gelatinase A (MMP2): structural implications for its function.";
RL FEBS LETT. 378:126-130(1996).
CC -I- CATALYTIC ACTIVITY: CLEAVAGE OF GELATIN TYPE I AND COLLAGEN TYPES
CC IV, V, VII, X. CLEAVES THE COLLAGEN-LIKE SEQUENCE PRO-GLN-GLY-|-
CC ILE-ALA-GLY-GLN.
CC -I- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.
CC -I- TISSUE SPECIFICITY: PRODUCED BY NORMAL SKIN FIBROBLASTS.
CC -I- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC -I- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE II-LIKE DOMAINS.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC
CC METALLOPROTEASE) ALSO KNOWN AS MATRIXIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J03210; G180671;
DR EMBL: M33789; G180601;
DR EMBL: M55593; G180616;
DR EMBL: M58552; G180616; JOINED.
DR EMBL: M55582; G180616; JOINED.
DR EMBL: M55583; G180616; JOINED.
DR EMBL: M55584; G180616; JOINED.
DR EMBL: M55585; G180616; JOINED.
DR EMBL: M55586; G180616; JOINED.
DR EMBL: M55587; G180616; JOINED.
DR EMBL: M55588; G180616; JOINED.
DR EMBL: M55589; G180616; JOINED.
DR EMBL: M55590; G180616; JOINED.
DR EMBL: M55591; G180616; JOINED.
DR EMBL: M55592; G180616; JOINED.
DR PIR: A28153; A28153.
DR PDB: 1RTG; 10-JUN-96.
DR PDB: 1GEN; 17-AUG-96.
DR MIM: 120360;
DR PROSITE: PS00023; FIBRONECTIN_2; 3.
DR PROSITE: PS00024; HEMOPEXIN; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00546; CYSTEINE_SWITCH; 1.
DR PFAM: PF00040; fn2; 3.
DR PFAM: PF00045; hemopexin; 4.
DR PFAM: PF00099; zn-protease; 1.
DR PFAM: PF00413; matrixin; 1.
DR HYDROLASE; METALLOPROTEASE; GLYCOPROTEIN; ZINC; ZYMOGEN; CALCIUM;

KW COLLAGEN DEGRADATION; EXTRACELLULAR MATRIX; DUPLICATION; SIGNAL;
KW 3D-STRUCTURE.
FT SIGNAL 1 29 POTENTIAL.
FT PROPEP 30 109 ACTIVATION PEPTIDE.
FT CHAIN 110 660 72 KD TYPE IV COLLAGENASE.
FT DOMAIN 100 107 AUTOINHIBITOR REGION.
FT DOMAIN 110 221 COLLAGENASE-LIKE.
FT DOMAIN 222 396 COLLAGEN-BINDING.
FT DOMAIN 397 465 COLLAGENASE-LIKE.
FT DOMAIN 466 660 HEMOPEXIN-LIKE.
FT DOMAIN 226 283 FIBRONECTIN TYPE-II.
FT DOMAIN 284 341 FIBRONECTIN TYPE-II.
FT DOMAIN 342 399 FIBRONECTIN TYPE-II.
FT METAL 403 403 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 404 404 BY SIMILARITY.
FT METAL 407 407 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 413 413 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 573 573 POTENTIAL.
FT CARBOHYD 642 642 POTENTIAL.
FT DISULFID 469 660
SQ SEQUENCE 660 AA; 73882 MW; 4372D672 CRC32;

Query Match 63.5%; Score 47; DB 1; Length 660;
Best Local Similarity 60.0%; Pred. No. 6.82e+00;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 497 VTNRDCKMGP 506
QY 355 VQRVPEMGP 364

Search completed: Thu Jul 8 18:48:53 1999
Job time : 7 secs.

WQREH (TW)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:49:13 1999; Maspar time 7.89 Seconds
69.161 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (355-364) from US09041236.pep (21 of 45)
Perfect Score: 74

Sequence: 1 VAQVPEPMGP 10

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 23.018; Variance 25.710; scale 0.895

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	74	100.0	666	4	SEMAPHORIN L.	2.92e-06
2	66	89.2	393	11	SEMAPHORIN L (FRAGMENT)	4.20e-04
3	52	70.3	546	2	MALTASE.	1.24e+00
4	52	70.3	1571	4	KIAA0434 (FRAGMENT).	1.24e+00
5	52	70.3	3942	11	BASSOON.	1.24e+00
6	51	68.9	202	13	RETINAL HOMEOBOX PROTE	2.10e+00
7	50	67.6	962	10	PULLULANASE-TYPE STARC	3.54e+00
8	49	66.2	118	1	FORMATE DEHYDROGENASE,	5.91e+00
9	49	66.2	166	13	TANDEMLY ARRANGED ENBR	5.91e+00
10	49	66.2	353	2	HYPOTHETICAL 40.0 KD P	5.91e+00
11	49	66.2	763	4	GASTRIN-BINDING PROTEIN	5.91e+00
12	49	66.2	763	6	GASTRIN-BINDING PROTEI	5.91e+00
13	49	66.2	763	6	GASTRIN-BINDING PROTEI	5.91e+00
14	49	66.2	773	6	EGF-2 BINDING PROTEIN	5.91e+00
15	48	64.9	770	13	CfR-ASSOCIATED PROTEIN	9.80e+00
16	48	64.9	1454	3	CHROMOSOME XII COSMID	9.80e+00
17	48	64.9	3938	11	BASSOON.	9.80e+00
18	47	63.5	172	14	REVERSE TRANSCRIPTASE-	1.61e+01
19	47	63.5	191	14	REVERSE TRANSCRIPTASE-	1.61e+01
20	47	63.5	335	7	MHC CLASS I PLA-A1 ALP	1.61e+01

21	47	63.5	382	2	006676	1.61e+01
22	47	63.5	450	1	027407	1.61e+01
23	47	63.5	675	13	Q08000	1.61e+01
24	47	63.5	774	5	Q20453	1.61e+01
25	47	63.5	866	5	P91426	1.61e+01
26	47	63.5	2530	3	Q92217	1.61e+01
27	46	62.2	127	2	054361	2.63e+01
28	46	62.2	173	5	044231	2.63e+01
29	46	62.2	330	10	082064	2.63e+01
30	46	62.2	545	2	005089	2.63e+01
31	46	62.2	561	3	074481	2.63e+01
32	46	62.2	701	2	086073	2.63e+01
33	46	62.2	1488	5	015849	2.63e+01
34	45	60.8	81	14	Q89002	4.26e+01
35	45	60.8	225	11	Q61167	4.26e+01
36	45	60.8	487	2	Q50901	4.26e+01
37	45	60.8	673	5	Q76549	4.26e+01
38	45	60.8	887	4	Q16507	4.26e+01
39	45	60.8	1436	11	Q35564	4.26e+01
40	45	60.8	1436	4	Q00197	4.26e+01
41	45	60.8	1439	4	P78399	4.26e+01
42	45	60.8	1442	4	Q92735	4.26e+01
43	45	60.8	1490	14	Q88791	4.26e+01
44	45	60.8	1612	5	Q27746	4.26e+01
45	45	60.8	2944	11	Q63870	4.26e+01

ALIGNMENTS

RESULT	1					
ID	075326	PRELIMINARY;	PRT;	666	AA.	
AC	075326;					
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)					
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)					
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)					
DE	SEMAPHORIN L.					
GN	SEMAL					
OS	HOMO SAPIENS (HUMAN).					
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;					
OC	CATARRHINI; HOMINIDAE; HOMO.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 98389619.					
RA	LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;					
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA					
RT	viruses.";					
RL	GENOMICS 51:340-350(1998).					
DR	EMBL; AF030698; G3523115; -					
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;					

Query Match 100.0%; Score 74; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 2.92e-06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	387	VAQVPEPMGP	396
Qy	355	VAQVPEPMGP	364

RESULT	2					
ID	088371	PRELIMINARY;	PRT;	393	AA.	
AC	088371;					
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)					
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)					
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)					
DE	SEMAPHORIN L (FRAGMENT).					
GN	SEMAL.					
OS	MUS MUSCULUS (MOUSE).					
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;					
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 98389619.					

RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RL VIRUSES";
RL GENOMICS 51:340-350(1998).
DR EMBL: AF030699; G3523117; -.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;
Query Match 89.2%; Score 66; DB 11; Length 393;
Best Local Similarity 100.0%; Pred. No. 4.20e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 385 VAQVEPMG 393
Qy 355 VAQVEPMG 363
RESULT 3 PRELIMINARY; PRT; 546 AA.
ID O53198
AC O53198;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MALTASE.
GN MYO008.27.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA MURPHY L., BROWN D., CHURCHER C.M., HARRIS D.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT leprae.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
DR EMBL: AL021246; E1237596; -.
SQ SEQUENCE 546 AA; 60443 MW; 417BBBD3 CRC32;
Query Match 70.3%; Score 52; DB 2; Length 546;
Best Local Similarity 75.0%; Pred. No. 1.24e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 5 QRPDPMPG 12
Qy 357 QRPDPMPG 364
RESULT 4 PRELIMINARY; PRT; 1571 AA.
ID O43161
AC O43161;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE KIA0434 (FRAGMENT).
GN KIA0434.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA ISHIIKAWA K., NAGASE T., NAKAJIMA D., SEKI N., OHIRA M., MIYAJIMA N.,
RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AB007894; D1024611; -.
FT NON_TER 1 1
SQ SEQUENCE 1571 AA; 170687 MW; 39C4CBBA CRC32;
Query Match 70.3%; Score 52; DB 4; Length 1571;
Best Local Similarity 66.7%; Pred. No. 1.24e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 896 SQRLEPLGP 904
Qy 356 AQRVEPMGP 364
RESULT 5 PRELIMINARY; PRT; 3942 AA.
ID O88737
AC O88737;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE BASSOON.
GN BASSOON.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-129 SVJ;
RA MEDLINE; 98345363.
RA DIECK S., SANMARTI-VILA L., LANGNAESE K., RICHTER K., KINDLER S.,
RA SOYKE A., WEX H., SMALLA K.H., KAMPE U., FRANZER J.T., STUMM M.,
RA GARNER C.C., GUNDELFINGER E.D.;
RT "Bassoon, a novel zinc-finger CAG/glutamine-repeat protein
RT selectively localized at the active zone of presynaptic nerve
RT terminals.";
RL J. CELL BIOL. 142:499-509(1998).
DR EMBL: Y17034; E1315350; -.
DR EMBL: Y17035; E1315350; JOINED.
DR EMBL: Y17036; E1315350; JOINED.
DR EMBL: Y17037; E1315350; JOINED.
DR EMBL: Y17038; E1315350; JOINED.
SQ SEQUENCE 3942 AA; 418739 MW; 9D6C5BC6 CRC32;
Query Match 70.3%; Score 52; DB 11; Length 3942;
Best Local Similarity 66.7%; Pred. No. 1.24e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 3260 SQRLEPLGP 3268
Qy 356 AQRVEPMGP 364
RESULT 6 PRELIMINARY; PRT; 202 AA.
ID O42356
AC O42356;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RETINAL HOMEOBOX PROTEIN (FRAGMENT).
GN RX1.
OS BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC CYPRINIDAE; RASBORINAE; DANIO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97320497.
RA MATHERS P.H., GRINBERG A., MAHON K.A., JAMRICH M.;
RT "The Rx homeobox gene is essential for vertebrate eye development.";

RL NATURE 387:603-607(1997).
CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AF001907; G2240026; -
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PFAM: PF00046; homeobox; 1.
KW HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN.
FT NON_TER 1
SQ SEQUENCE 202 AA; 23261 MW; 957612C6 CRC32;

Query Match 68.9%; Score 51; DB 13; Length 202;
Best Local Similarity 60.0%; Pred. No. 2.10e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 146 MAQSNOPMGP 155
QY 355 VAQVRPMGP 364
|||||

RESULT 7
ID O81638 PRELIMINARY; PRT; 962 AA.
AC O81638;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PULLULANASE-TYPE STARCH DEBRANCHING ENZYME.
GN ZPUL.
OS ZEA MAYS (MAIZE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; ZEA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ENDOSPERM;
RA BEATTY M.K., RAHMAN A., MYERS A.M., JAMES M.G.;
RT "Genetic and biochemical characterization of zpul, a pullulanase-type
starch debranching enzyme from maize."
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF080567; G3411266; -
SQ SEQUENCE 962 AA; 106239 MW; 0FICA2BD CRC32;

Query Match 67.6%; Score 50; DB 10; Length 962;
Best Local Similarity 70.0%; Pred. No. 3.54e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 48 VAQVRVPAP 57
QY 355 VAQVRPMGP 364
|||||

RESULT 8
ID O34188 PRELIMINARY; PRT; 118 AA.
AC O34188;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE FORMATE DEHYDROGENASE, ALPHA SUBUNIT (FRAGMENT).
OS HALOBACTERIUM VOLCANII (HALOFERAX VOLCANII).
OG PLASMID PHV3.
OC ARCHAEA; EURYARCHAEOTA; HALOBACTERIALES; HALOBACTERIACEAE; HALOFERAX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DS2;
RA FARAHANI R., IMBEAULT J.-C., ST JEAN A., CHAN C.C.-Y., ALLARD G.,
RA CHARLEBOIS R.L.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U95376; G2459739; -
KW PLASMD.
FT NON_TER 1
SQ SEQUENCE 118 AA; 12479 MW; 2E681822 CRC32;

Query Match 66.2%; Score 49; DB 1; Length 118;
Best Local Similarity 85.7%; Pred. No. 5.91e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 90 RVEPLGP 96
QY 358 RVEPMGP 364
|||||

RESULT 9
ID Q91873 PRELIMINARY; PRT; 166 AA.
AC Q91873;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TANDEMLY ARRANGED EMBRYONIC UL SNRNA GENES UIA/UIB.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85284944.
RA KROL A., LUND E., DAHLBERG J.E.;
RT "The two embryonic ul RNA genes of Xenopus laevis have both common
RT and gene-specific transcription signals."
RL ENBO J. 4:1529-1535(1985).
DR EMBL: X02585; G65183; -
SQ SEQUENCE 166 AA; 18233 MW; 4E9601C1 CRC32;

Query Match 66.2%; Score 49; DB 13; Length 166;
Best Local Similarity 62.5%; Pred. No. 5.91e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 52 ERVOPLGP 59
QY 357 QRVPMGP 364
|||||

RESULT 10
ID O84359 PRELIMINARY; PRT; 353 AA.
AC O84359;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 40.0 KD PROTEIN.
GN CT355.
OS CHLAMYDIA TRACHOMATIS.
OC BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis."
RL SCIENCE 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE001308; G3328776; -
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 353 AA; 39969 MW; 8DB64D55 CRC32;

Query Match 66.2%; Score 49; DB 2; Length 353;
Best Local Similarity 87.5%; Pred. No. 5.91e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 156 VAORVEGM 163
QY 355 VAQVRPMGP 362
|||||

```

RESULT 11
ID Q16679 PRELIMINARY; PRT; 763 AA.
AC Q16679;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GASTRIN-BINDING PROTEIN.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95002180.
RA ZHANG Q.X., BALDWIN G.S.;
RT "Structures of the human cDNA and gene encoding the 78 kDa
RT gastrin-binding protein and of a related pseudogene.";
RL BIOCHIM. BIOPHYS. ACTA 1219:567-573(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA BALDWIN G.S.;
RL COMP. BIOCHEM. PHYSIOL. 104:55-61(1993).
DR EMBL; U04627; G595267; -.
DR PROSITE; PS00067; 3HCDH; 1.
DR PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
DR PFAM; PF00378; ECH; 1.
DR PFAM; PF00725; 3HCDH; 1.
SQ SEQUENCE 763 AA; 83041 MW; 9245DFC9 CRC32;

Query Match 66.2%; Score 49; DB 4; Length 763;
Best Local Similarity 70.0%; Pred. No. 5.91e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 218 VDQVLEPLGP 227
QY 355 VAQRVEPMGP 364

RESULT 12
ID Q62831 PRELIMINARY; PRT; 763 AA.
AC Q62831;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE GASTRIN BINDING PROTEIN-LIKE PRECURSOR (FRAGMENT).
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RA LEMAITRE G., CAILLERET K., CHEVET E., DAHAN S., BERGERON J.J.,
RA KATINKA M.D.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AJ222637; E1285541; -.
DR PROSITE; PS00067; 3HCDH; 1.
DR PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
KW SIGNAL.
FT SIGNAL. 1 36 POTENTIAL.
FT CHAIN 37 >763 POTENTIAL.
FT NON_TER 763 763
SQ SEQUENCE 763 AA; 83291 MW; EAD83B3F CRC32;

Query Match 66.2%; Score 49; DB 6; Length 763;
Best Local Similarity 70.0%; Pred. No. 5.91e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 218 VDQVLEPLGP 227
QY 355 VAQRVEPMGP 364

RESULT 13

```

```

ID Q29554 PRELIMINARY; PRT; 763 AA.
AC Q29554;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GASTRIN-BINDING PROTEIN.
GN LCHYD-HAD.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94002212.
RA MANTAMADIOTIS T., SOBIESZCZUK P., WEINSTOCK J., BALDWIN G.S.;
RT "Nucleotide sequence encoding a novel member of the
RT hydratase/dehydrogenase family.";
RL BIOCHIM. BIOPHYS. ACTA 1170:211-215(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93193423.
RA BALDWIN G.S.;
RT "Comparison of sequences of the 78 kDa gastrin-binding protein and
RT some enzymes involved in fatty acid oxidation.";
RL COMP. BIOCHEM. PHYSIOL. B, COMP. BIOCHEM. 104:55-61(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93290643.
RA BALDWIN G.S., CASEY A., WEINSTOCK J.;
RT "Partial structure of the gene encoding the 78 kDa gastrin binding
RT protein excludes a close relationship with the peroxisomal
RT trifunctional enzyme.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 193:560-564(1993).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94128080.
RA YANG S.Y., HE X.Y., STYLES J., LUO M.J., SCHULZ H., ELZINGA M.;
RT "Primary structure of the large subunit of trifunctional
RT beta-oxidation complex from pig heart mitochondria.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 198:431-437(1994).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95187623.
RA YANG S.Y.;
RT "The large subunit of the pig heart mitochondrial membrane-bound
RT beta-oxidation complex is a long-chain enoyl-CoA hydratase:
RT 3-hydroxyacyl-CoA dehydrogenase bifunctional enzyme.";
RL COMP. BIOCHEM. PHYSIOL. BIOCHEM. MOL. BIOL. 109:557-566(1994).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95187623.
RA YANG S.Y.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; L12581; G433066; -.
DR EMBL; AF028609; G2599088; -.
DR PROSITE; PS00067; 3HCDH; 1.
DR PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
DR PFAM; PF00378; ECH; 1.
DR PFAM; PF00725; 3HCDH; 1.
KW LYASE.
SQ SEQUENCE 763 AA; 83106 MW; 41CDEB71 CRC32;

Query Match 66.2%; Score 49; DB 6; Length 763;
Best Local Similarity 70.0%; Pred. No. 5.91e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 218 VDQVLEPLGP 227
QY 355 VAQRVEPMGP 364

```

```

RESULT 15
ID Q57660 PRELIMINARY; PRT; 770 AA.
AC Q57660;
DT 01-JUN-1998 (TREMBUREL 06, CREATED)
DT 01-JUN-1998 (TREMBUREL 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBUREL 07, LAST ANNOTATION UPDATE)
DT CFR-ASSOCIATED PROTEIN P70.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ZUBER M.E., BURRUS L.W., ZHOU Z., OLWIN B.B.;
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL: U83249; G2737971: -
DR PROSITE; PS00067; 3HCDH; 1.
DR PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
SQ SEQUENCE 770 AA; 83186 MW; FE18C090 CRC32;

Query Match 64.9%; Score 48; DB 13; Length 770;
Best Local Similarity 60.0%; Pred. No. 9.80e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 225 VDQLVDPLGP 234
| | : | : |
Qy 355 VAQRVPMGP 364

Search completed: Thu Jul 8 18:49:25 1999
Job time : 12 secs.

```

THIS PAGE BLANK (USPTO)

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:52:21 1999; Maspar time 11.15 Seconds

Tabular output not generated. 19.076 Million cell updates/sec

Title: >US-09-041-236-2
Description: (365-374) from US09041236.pep (22 of 45)
Perfect Score: 79
Sequence: 1 LKTPLFHSKY 10

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 17.608; Variance 49.597; scale 0.355

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	67.1	481	36	HSV-2 strain SB5 Cont	4.14e+01
2	53	67.1	511	36	HSV-2 strain SB5 Cont	4.14e+01
3	53	67.1	511	36	HSV-2 strain SB5 Cont	4.14e+01
4	53	67.1	572	31	W51159	4.14e+01
5	51	64.6	32	39	Rhodospiridium torulo	6.97e+01
6	51	64.6	278	33	Human zcytor5 epitope	6.97e+01
7	51	64.6	303	39	Human zcytor5 variant	6.97e+01
8	51	64.6	303	39	Human zcytor5 variant	6.97e+01
9	51	64.6	350	33	Human zcytor5 variant	6.97e+01
10	51	64.6	385	39	Human zcytor5 variant	6.97e+01
11	51	64.6	385	39	Human zcytor5 variant	6.97e+01
12	51	64.6	388	39	Human zcytor5 variant	6.97e+01
13	51	64.6	389	39	Human zcytor5 variant	6.97e+01
14	51	64.6	389	39	Human zcytor5 variant	6.97e+01
15	51	64.6	389	39	Human zcytor5 variant	6.97e+01
16	51	64.6	389	39	Human zcytor5 variant	6.97e+01

17	51	64.6	389	39	W70848	Human zcytor5 variant	6.97e+01
18	51	64.6	389	39	W70851	Human zcytor5 variant	6.97e+01
19	51	64.6	389	39	W70850	Human zcytor5 variant	6.97e+01
20	51	64.6	389	39	W70852	Human zcytor5 variant	6.97e+01
21	51	64.6	389	39	W70849	Human zcytor5 variant	6.97e+01
22	51	64.6	392	39	W70840	Human zcytor5 variant	6.97e+01
23	51	64.6	408	34	W59805	Amino acid sequence o	6.97e+01
24	51	64.6	413	33	W55011	Novel haemopoietin re	6.97e+01
25	51	64.6	422	39	W70860	Human zcytor5 protein	6.97e+01
26	51	64.6	425	39	W70861	Allelic variant of hu	6.97e+01
27	51	64.6	425	39	W70862	Rat zcytor5 protein.	6.97e+01
28	51	64.6	425	34	W59804	Nucleotide sequence o	6.97e+01
29	51	64.6	425	33	W55012	Novel haemopoietin re	6.97e+01
30	49	62.0	332	32	W60768	Murine radical fringe	1.17e+02
31	49	62.0	372	23	W22065	Chick fringe A (radic	1.17e+02
32	49	62.0	430	33	W63843	S. aureus asparaginy	1.17e+02
33	49	62.0	432	33	W63842	S. aureus asparaginy	1.17e+02
34	49	62.0	432	23	W21894	AsparaginyI-trNA synt	1.17e+02
35	49	62.0	510	19	W04334	Human lymphoid cell a	1.17e+02
36	49	62.0	510	21	W04264	Human CD39 protein.	1.17e+02
37	48	60.8	551	24	W27150	HMG-CoA reductase deg	1.50e+02
38	47	59.5	279	8	R42506	Protein encoded by Ro	1.93e+02
39	47	59.5	308	2	P94810	Sequence of horseradi	1.93e+02
40	47	59.5	308	1	P81167	Horseradish peroxidase	1.93e+02
41	47	59.5	309	1	P95200	Horseradish peroxidase	1.93e+02
42	47	59.5	342	1	P95203	Horseradish peroxidase	1.93e+02
43	47	59.5	349	6	R32760	HRP/IGF-II fusion pro	1.93e+02
44	46	58.2	666	34	W62829	Macadamia integrifoli	2.48e+02
45	46	58.2	732	19	W06402	Kell protein sequence	2.48e+02

ALIGNMENTS

RESULT 1

ID W72161 standard; Protein; 481 AA.
AC W72161:
DT 08-JAN-1999 (first entry)
DE HSV-2 strain SB5 Contig ID 12 ORF#3 protein.
KW HSV-2 strain SB5; Immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor.
OS Herpes simplex virus type 2.
PN W09820016-A1.
PD 14-MAY-1998.
PF 31-OCT-1997; U20016.
PR 01-JUN-1997; US-049018.
PR 04-NOV-1996; US-030279.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,
PI Esser KM, Leary JJ;
DR WPI; 98-286847/25.
DR N-PSDB; V62175.
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal
PS Claim 10; Page 105; 748pp; English.
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 12.
CC The proteins can be used for the treatment or prevention of disease, to
CC induce an immunological response in a mammal or to identify inhibitors,
CC activators or novel antivirals. Antagonists of the proteins can be used
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC it can also be used to induce an immunological response in a mammal.
SQ Sequence 481 AA;

Query Match 67.1%; Score 53; DB 36; Length 481;
Best Local Similarity 66.7%; Pred. No. 4.14e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 473 lrlplfhpk 481
l: lll:l
Qy 365 LKTPLFHSK 373

RESULT 2
ID W72141 standard; Protein; 511 AA.
AC W72141; 1998 (first entry)
DT 23-DEC-1998 (first entry)
DE HSV-2 strain SB5 Contig ID 3 ORF#3 protein.
KW HSV-2 strain SB5; immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor.
OS Herpes simplex virus type 2.
PN W09820016-A1.
PD 14-MAY-1998.
PF 31-OCT-1997; U20016.
PR 09-JUN-1997; US-049018.
PR 04-NOV-1996; US-030279.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,
PI Esser KM, Leary JJ;
DR WPI; 98-286847/25.
DR N-PSDB; V62154.
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal.
PS Claim 10; Page 98; 748pp; English.
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 3.
CC The proteins can be used for the treatment or prevention of disease, to
CC induce an immunological response in a mammal or to identify inhibitors,
CC activators or novel antivirals. Antagonists of the proteins can be used
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC it can also be used to induce an immunological response in a mammal.
SQ Sequence 511 AA;

Query Match 67.1%; Score 53; DB 36; Length 511;
Best Local Similarity 66.7%; Pred. No. 4.14e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 503 lrlplfhpk 511
|:||||:|
QY 365 LKTPLEFHSK 373

RESULT 3
ID W72073 standard; Protein; 511 AA.
AC W72073;
DT 18-DEC-1998 (first entry)
DE HSV-2 strain SB5 Contig ID 96 ORF#1 protein.
KW HSV-2 strain SB5; immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor.
OS Herpes simplex virus type 2.
PN W09820016-A1.
PD 14-MAY-1998.
PF 31-OCT-1997; U20016.
PR 09-JUN-1997; US-049018.
PR 04-NOV-1996; US-030279.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,
PI Esser KM, Leary JJ;
DR WPI; 98-286847/25.
DR N-PSDB; V62150.
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal.
PS Claim 10; Page 69; 748pp; English.
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 96.
CC The proteins can be used for the treatment or prevention of disease, to
CC induce an immunological response in a mammal or to identify inhibitors,
CC activators or novel antivirals. Antagonists of the proteins can be used
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC it can also be used to induce an immunological response in a mammal.
SQ Sequence 511 AA;

Query Match 67.1%; Score 53; DB 36; Length 511;
Best Local Similarity 66.7%; Pred. No. 4.14e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 503 lrlplfhpk 511
|:||||:|
QY 365 LKTPLEFHSK 373

Query Match 67.1%; Score 53; DB 36; Length 511;
Best Local Similarity 66.7%; Pred. No. 4.14e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 503 lrlplfhpk 511
|:||||:|
QY 365 LKTPLEFHSK 373

RESULT 4
ID W54159 standard; Protein; 572 AA.
AC W54159;
DT 17-AUG-1998 (first entry)
DE Rhodospiridium toruloides cephalosporin esterase.
KW Cephalosporin esterase; antibiotic.
OS Rhodospiridium toruloides strain ATCC 10657.
FH Key Location/Qualifiers
FT Peptide 1..28
FT /label= Sig_peptide
FT Protein 29..572
FT /label= Mat_protein
FT /notes= "(Claim 14)"
FT W09812345-A1.
PN 26-MAR-1998.
PD 11-SEP-1997; U16193.
PR 18-SEP-1996; US-026929.
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Burnett WV, Politino M, Romancik G, Tonzi SM, Usher JJ;
DR WPI; 98-230321/20.
DR N-PSDB; V21460-61.
PT Isolated cephalosporin esterase gene - obtained from Rhodospiridium
PT toruloides, used to obtain products which can be used for hydrolysis
PT of cephalosporins
PS Claim 13; Fig 5; 38pp; English.
CC This protein comprises Rhodospiridium toruloides ATCC 10657
CC cephalosporium esterase, an enzyme that catalyses the hydrolysis of
CC the 3' acetyl groups of cephalosporins. The enzyme, especially the
CC mature polypeptide, can be purified from cultures of R. toruloides
CC or expressed in claimed host cells (preferably Escherichia coli, R.
CC toruloides, Cephalosporium acremonium or Penicillium chrysogenum)
CC transformed with a vector comprising cephalosporin esterase cDNA or
CC genomic DNA (see V21460-61). The isolated polypeptides, even if
CC catalytically inactive, can also be used for the production of
CC antibodies for use in detection and purification.
SQ Sequence 572 AA;

Query Match 67.1%; Score 53; DB 31; Length 572;
Best Local Similarity 75.0%; Pred. No. 4.14e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 256 lkkplfha 263
|||
QY 365 LKTPLEFHS 372

RESULT 5
ID W70857 standard; Peptide; 32 AA.
AC W70857;
DT 17-MAR-1999 (first entry)
DE Human Zcytor5 epitope-bearing region.
KW Zcytor5; cytokinin-like receptor; down-regulation; growth factor;
KW maintenance factor; thymoid; heart; skeletal muscle; cardiostrophin-1;
KW cardiac pathology; heart enlargement; Zcytor5 ligand; epitope.
OS Homo sapiens.
PN W09849307-A1.
PD 05-NOV-1998.
PF 01-MAY-1998; U08865.
PR 13-FEB-1998; US-074721.
PR 01-MAY-1997; US-045287.
PR 01-MAY-1997; US-850030.
PR 13-FEB-1998; US-023890.
PA (ZYMO) ZYMOGENETICS INC.

PI Adams RL, Foster DC, Gilbert T, Jelmeberg AC, Lehner JM,
 PI Lok S, Presnell SR, Whitmore TE;
 DR WPI: 99-034662/03.
 PT New mammalian cytokinin-like receptor Zcytor5 - useful for, e.g.
 PT down-regulating Zcytor5 natural ligands or detecting cardiostrophin-1
 PT in blood
 PS Claim 1: Page 102; 55pp; English.
 CC The present sequence represents a Zcytor5 epitope-bearing region. Zcytor5
 CC is a cytokinin-like receptor. Soluble Zcytor5 may be administered to
 CC down-regulate the effects of a growth and/or maintenance factor in
 CC thyroid, heart, and skeletal muscle for example to lessen the effect
 CC of cardiostrophin-1 on cardiac pathologies, so preventing heart
 CC enlargement. Zcytor5 could be used to detect cardiostrophin-1 in the
 CC blood, and to discover other possible Zcytor5 ligands. A probe
 CC comprising Zcytor5 DNA or RNA can be used to determine the presence
 CC and integrity of the Zcytor5 gene on chromosome 19. Antibodies and the
 CC anti-idiotypic antibody could be used to purify Zcytor5 and
 CC therapeutically to modify Zcytor5 ligand effects.
 SQ Sequence 32 AA;
 Query Match 64.6%; Score 51; DB 39; Length 32;
 Best Local Similarity 60.0%; Pred. No. 6.97e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 5 lkdfifgaky 14
 QY 365 LKTPLFHSKY 374
 RESULT 6
 ID W5014 standard; Protein: 278 AA.
 DT 02-OCT-1998 (first entry)
 DE Protein sequence of products generated by 5N race of brain cDNA.
 KW Haemopoietin receptor; cell proliferation; cell differentiation; cancer;
 KW cell survival; therapeutic; neuronal proliferation; drug screening;
 KW Mouse.
 OS Mus sp.
 PN W09811225-A2.
 PD 19-MAR-1998.
 PF 11-SEP-1997; G02479.
 PR 11-SEP-1996; AU-002246.
 PA (AMRA-) AMRAD OPERATIONS PTY LTD.
 PA (DZIE/) DZIEGLEWSKA H E.
 PI Alexander W, Fabri L, Farley A, Hilton DJ, Kikuchi Y,
 PI Kojima T, Maeda M, Nash A, Nicola NA, Rakar S, Willson T,
 PI Zhang J;
 DR WPI: 98-260970/23.
 DR N-PSDB; V27143.
 PT New isolated haemopoietin receptor - used for developing products
 PT for modulating proliferation, differentiation and survival of cells,
 PT e.g. neuronal cells
 PS Claim 17; Page 93-95; 182pp; English.
 CC The protein sequence was generated by a 5N RACE of brain cDNA using
 CC NR6 specific primers. NR6 is a novel Haemopoietin receptor (HR).
 CC Interaction between the novel HR and a ligand facilitates proliferation,
 CC differentiation and survival of a wide variety of cells. The HR and it's
 CC derivatives can be used for modulating the activity of the receptors e.g.
 CC to regulate development, maintenance or regeneration in an array of
 CC different cells and tissues in vitro and in vivo. They can be present in
 CC therapeutics used for modulating neuronal proliferation, differentiation
 CC and survival. The products can also be used for detection and diagnosis,
 CC e.g. for cancers or predisposition to cancers, or for drug screening.
 SQ Sequence 278 AA;
 Query Match 64.6%; Score 51; DB 33; Length 278;
 Best Local Similarity 60.0%; Pred. No. 6.97e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 216 lkdfifgaky 225
 QY 365 LKTPLFHSKY 374

RESULT 7
 ID W70843 standard; Protein: 303 AA.
 AC W70843;
 DT 17-MAR-1999 (first entry)
 DE Human Zcytor5 variant.
 KW Zcytor5; cytokinin-like receptor; down-regulation; growth factor;
 KW maintenance factor; thyroid; heart; skeletal muscle; cardiostrophin-1;
 KW cardiac pathology; heart enlargement; Zcytor5 ligand; variant.
 OS Homo sapiens.
 PN W09849307-A1.
 PD 05-NOV-1998.
 PF 01-MAY-1998; U08865.
 PR 13-FEB-1998; US-074721.
 PR 01-MAY-1997; US-045287.
 PR 01-MAY-1997; US-850030.
 PR 13-FEB-1998; US-023890.
 PA (ZYMO) ZYMOGENETICS INC.
 PI Adams RL, Foster DC, Gilbert T, Jelmeberg AC, Lehner JM,
 PI Lok S, Presnell SR, Whitmore TE;
 DR WPI: 99-034662/03.
 PT New mammalian cytokinin-like receptor Zcytor5 - useful for, e.g.
 PT down-regulating Zcytor5 natural ligands or detecting cardiostrophin-1
 PT in blood
 PS Claim 1: Page 86; 55pp; English.
 CC The present sequence represents a Zcytor5 variant protein. Zcytor5
 CC is a cytokinin-like receptor. Soluble Zcytor5 may be administered to
 CC down-regulate the effects of a growth and/or maintenance factor in
 CC thyroid, heart, and skeletal muscle for example to lessen the effect
 CC of cardiostrophin-1 on cardiac pathologies, so preventing heart
 CC enlargement. Zcytor5 could be used to detect cardiostrophin-1 in the
 CC blood, and to discover other possible Zcytor5 ligands. A probe
 CC comprising Zcytor5 DNA or RNA can be used to determine the presence
 CC and integrity of the Zcytor5 gene on chromosome 19. Antibodies and the
 CC anti-idiotypic antibody could be used to purify Zcytor5 and
 CC therapeutically to modify Zcytor5 ligand effects.
 SQ Sequence 303 AA;
 Query Match 64.6%; Score 51; DB 39; Length 303;
 Best Local Similarity 60.0%; Pred. No. 6.97e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 226 lkdfifgaky 235
 QY 365 LKTPLFHSKY 374

RESULT 8
 ID W70845 standard; Protein: 303 AA.
 AC W70845;
 DT 17-MAR-1999 (first entry)
 DE Human Zcytor5 variant.
 KW Zcytor5; cytokinin-like receptor; down-regulation; growth factor;
 KW maintenance factor; thyroid; heart; skeletal muscle; cardiostrophin-1;
 KW cardiac pathology; heart enlargement; Zcytor5 ligand; variant.
 OS Homo sapiens.
 PN W09849307-A1.
 PD 05-NOV-1998.
 PF 01-MAY-1998; U08865.
 PR 13-FEB-1998; US-074721.
 PR 01-MAY-1997; US-045287.
 PR 01-MAY-1997; US-850030.
 PR 13-FEB-1998; US-023890.
 PA (ZYMO) ZYMOGENETICS INC.
 PI Adams RL, Foster DC, Gilbert T, Jelmeberg AC, Lehner JM,
 PI Lok S, Presnell SR, Whitmore TE;
 DR WPI: 99-034662/03.
 PT New mammalian cytokinin-like receptor Zcytor5 - useful for, e.g.
 PT down-regulating Zcytor5 natural ligands or detecting cardiostrophin-1
 PT in blood
 PS Claim 1: Page 88-89; 55pp; English.
 CC The present sequence represents a Zcytor5 variant protein. Zcytor5
 CC is a cytokinin-like receptor. Soluble Zcytor5 may be administered to

CC down-regulate the effects of a growth and/or maintenance factor in
 CC thyroid, heart, and skeletal muscle for example to lessen the effect
 CC of cardiostrophin-1 on cardiac pathologies, so preventing heart
 CC enlargement. Zcytor5 could be used to detect cardiostrophin-1 in the
 CC blood, and to discover other possible Zcytor5 ligands. A probe
 CC comprising Zcytor5 DNA or RNA can be used to determine the presence
 CC and integrity of the Zcytor5 gene on chromosome 19. Antibodies and the
 CC anti-idiotypic antibody could be used to purify Zcytor5 and
 CC therapeutically to modify Zcytor5 ligand effects.
 SQ Sequence 303 AA;

Query Match 64.6%; Score 51; DB 39; Length 303;
 Best Local Similarity 60.0%; Pred. No. 6.97e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 226 lkdflfqaky 235
 || ||::||
 QY 365 LKTPLFHSKY 374

RESULT 9

ID W55015 standard; Protein; 350 AA.

AC W55015;

DT 29-SEP-1998 (first entry)

DE Amino acid sequence of clone HFK-66 encoding human NR6.

KW Haemopoietin receptor; cell proliferation; cell differentiation; cancer;

KW cell survival; therapeutic; neuronal proliferation; drug screening;

KW Human.

OS Homo sapiens.

PN WO9811225-A2.

PD 19-MAR-1998.

PF 11-SEP-1997; G02479.

PR 11-SEP-1996; AU-002246.

PA (AMRA-) AMRAD OPERATIONS PTY LTD.

PA (DZIE/) DZIELEWSKA H E.

PI Alexander W, Fabri L, Farley A, Hilton DJ, Kikuchi Y,

PI Kojima T, Maeda M, Nash A, Nicola NA, Rakar S, Willson T,

PI Zhang J;

DR WPI: 98-260970/23.

DR N-PSDB: V27144.

PT New isolated haemopoietin receptor - used for developing products

PT for modulating proliferation, differentiation and survival of cells,

PT e.g. neuronal cells

PS Claim 18: Page 102-104; 182pp; English.

CC The NR6 protein is a novel Haemopoietin receptor (HR). Interaction

CC between the novel HR and a ligand facilitates proliferation,

CC differentiation and survival of a wide variety of cells. The HR and it's

CC derivatives can be used for modulating the activity of the receptors e.g.

CC to regulate development, maintenance or regeneration in an array of

CC different cells and tissues in vitro and in vivo. They can be present in

CC therapeutics used for modulating neuronal proliferation, differentiation

CC and survival. The products can also be used for detection and diagnosis,

CC e.g. for cancers or predisposition to cancers, or for drug screening.

SQ Sequence 350 AA;

Query Match 64.6%; Score 51; DB 33; Length 350;
 Best Local Similarity 60.0%; Pred. No. 6.97e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 187 lkdflfqaky 196
 || ||::||
 QY 365 LKTPLFHSKY 374

RESULT 10

ID W70842 standard; Protein; 385 AA.

AC W70842;

DT 17-MAR-1999 (first entry)

DE Human Zcytor5 variant.

KW Zcytor5; cytokinin-like receptor; down-regulation; growth factor;

KW maintenance factor; thyroid; heart; skeletal muscle; cardiostrophin-1;

KW cardiac pathology; heart enlargement; Zcytor5 ligand; variant.

OS Homo sapiens.

FH Key Location/Qualifiers
 FT Misc_difference 42 /note= "not specified"
 FT PN WO9849307-A1.

PD 05-NOV-1998.

PF 01-MAY-1998; U08865.

PR 13-FEB-1998; US-074721.

PR 01-MAY-1997; US-045287.

PR 01-MAY-1997; US-850030.

PR 13-FEB-1998; US-023890.

PA (ZYMO) ZYMOGENETICS INC.

PI Adams RL, Foster DC, Gilbert T, Jelmsberg AC, Lehner JM,

PI Lok S, Presnell SR, Whitmore TE;

DR WPI: 99-034662/03.

PT New mammalian cytokinin-like receptor Zcytor5 - useful for, e.g.
 PT down-regulating Zcytor5 natural ligands or detecting cardiostrophin-1
 PT in blood

PS Claim 1; Page 84-85; 55pp; English.

CC The present sequence represents a Zcytor5 variant protein. Zcytor5
 CC is a cytokinin-like receptor. Soluble Zcytor5 may be administered to
 CC down-regulate the effects of a growth and/or maintenance factor in
 CC thyroid, heart, and skeletal muscle for example to lessen the effect
 CC of cardiostrophin-1 on cardiac pathologies, so preventing heart
 CC enlargement. Zcytor5 could be used to detect cardiostrophin-1 in the
 CC blood, and to discover other possible Zcytor5 ligands. A probe
 CC comprising Zcytor5 DNA or RNA can be used to determine the presence
 CC and integrity of the Zcytor5 gene on chromosome 19. Antibodies and the
 CC anti-idiotypic antibody could be used to purify Zcytor5 and
 CC therapeutically to modify Zcytor5 ligand effects.

SQ Sequence 385 AA;

Query Match 64.6%; Score 51; DB 39; Length 385;
 Best Local Similarity 60.0%; Pred. No. 6.97e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 226 lkdflfqaky 235
 || ||::||
 QY 365 LKTPLFHSKY 374

RESULT 11

ID W70841 standard; Protein; 385 AA.

AC W70841;

DT 17-MAR-1999 (first entry)

DE Human Zcytor5 variant.

DZ Zcytor5; cytokinin-like receptor; down-regulation; growth factor;

KW maintenance factor; thyroid; heart; skeletal muscle; cardiostrophin-1;

KW cardiac pathology; heart enlargement; Zcytor5 ligand; variant.

OS Homo sapiens.

PN WO9849307-A1.

PD 05-NOV-1998.

PF 01-MAY-1998; U08865.

PR 13-FEB-1998; US-074721.

PR 01-MAY-1997; US-045287.

PR 01-MAY-1997; US-850030.

PR 13-FEB-1998; US-023890.

PA (ZYMO) ZYMOGENETICS INC.

PI Adams RL, Foster DC, Gilbert T, Jelmsberg AC, Lehner JM,

PI Lok S, Presnell SR, Whitmore TE;

DR WPI: 99-034662/03.

PT New mammalian cytokinin-like receptor Zcytor5 - useful for, e.g.
 PT down-regulating Zcytor5 natural ligands or detecting cardiostrophin-1
 PT in blood

PS Claim 1; Page 83-84; 55pp; English.

CC The present sequence represents a Zcytor5 variant protein. Zcytor5
 CC is a cytokinin-like receptor. Soluble Zcytor5 may be administered to
 CC down-regulate the effects of a growth and/or maintenance factor in
 CC thyroid, heart, and skeletal muscle for example to lessen the effect
 CC of cardiostrophin-1 on cardiac pathologies, so preventing heart
 CC enlargement. Zcytor5 could be used to detect cardiostrophin-1 in the
 CC blood, and to discover other possible Zcytor5 ligands. A probe
 CC comprising Zcytor5 DNA or RNA can be used to determine the presence
 CC and integrity of the Zcytor5 gene on chromosome 19. Antibodies and the

CC anti-idiotypic antibody could be used to purify zcytor5 and
 CC therapeutically to modify zcytor5 ligand effects.
 SQ Sequence 385 AA;

Query Match 64.6%; Score 51; DB 39; Length 385;
 Best Local Similarity 60.0%; Pred. No. 6.97e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 226 lkdfifqaky 235
 QY 365 LKTPLFHSKY 374
 || ||:|

RESULT 12

ID W70839 standard; Protein: 388 AA.
 AC W70839;
 DT 17-MAR-1999 (first entry)
 DE Human zcytor5 variant.
 KW zcytor5; cytokinin-like receptor; down-regulation; growth factor;
 KW maintenance factor; thyroid; heart; skeletal muscle; cardiostrophin-1;
 KW cardiac pathology; heart enlargement; zcytor5 ligand; variant.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 45
 FT Note= "not specified"

PN W09849307-Al.
 PD 05-NOV-1998.
 PF 01-MAY-1998; U08865.
 PR 13-FEB-1998; US-074721.
 PR 01-MAY-1997; US-045287.
 PR 01-MAY-1997; US-850030.
 PR 13-FEB-1998; US-023890.
 PA (ZYMO) ZYMOGENETICS INC.
 PI Adams RL, Foster DC, Gilbert T, Jelmberg AC, Lehner JM,
 PI Lok S, Presnell SR, Whitmore TE;
 DR WPI: 99-034662/03.
 PT New mammalian cytokinin-like receptor zcytor5 - useful for, e.g.
 PT down-regulating zcytor5 natural ligands or detecting cardiostrophin-1
 PT in blood
 PS Claim 1; Page 80-81; 55pp; English.

CC The present sequence represents a zcytor5 variant protein. zcytor5
 CC is a cytokinin-like receptor. Soluble zcytor5 may be administered to
 CC down-regulate the effects of a growth and/or maintenance factor in
 CC thyroid, heart, and skeletal muscle for example to lessen the effect
 CC of cardiostrophin-1 on cardiac pathologies, so preventing heart
 CC enlargement. zcytor5 could be used to detect cardiostrophin-1 in the
 CC blood, and to discover other possible zcytor5 ligands. A probe
 CC comprising zcytor5 DNA or RNA can be used to determine the presence
 CC and integrity of the zcytor5 gene on chromosome 19. Antibodies and the
 CC anti-idiotypic antibody could be used to purify zcytor5 and
 CC therapeutically to modify zcytor5 ligand effects.
 SQ Sequence 388 AA;

Query Match 64.6%; Score 51; DB 39; Length 388;
 Best Local Similarity 60.0%; Pred. No. 6.97e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 229 lkdfifqaky 238
 QY 365 LKTPLFHSKY 374
 || ||:|

RESULT 13

ID W70844 standard; Protein: 389 AA.
 AC W70844;
 DT 17-MAR-1999 (first entry)
 DE Human zcytor5 variant.
 KW zcytor5; cytokinin-like receptor; down-regulation; growth factor;
 KW maintenance factor; thyroid; heart; skeletal muscle; cardiostrophin-1;
 KW cardiac pathology; heart enlargement; zcytor5 ligand; variant.
 OS Homo sapiens.
 PN W09849307-Al.
 PD 05-NOV-1998.

PF 01-MAY-1998; U08865.
 PR 13-FEB-1998; US-074721.
 PR 01-MAY-1997; US-045287.
 PR 01-MAY-1997; US-850030.
 PR 13-FEB-1998; US-023890.
 PA (ZYMO) ZYMOGENETICS INC.
 PI Adams RL, Foster DC, Gilbert T, Jelmberg AC, Lehner JM,
 PI Lok S, Presnell SR, Whitmore TE;
 DR WPI: 99-034662/03.
 PT New mammalian cytokinin-like receptor zcytor5 - useful for, e.g.
 PT down-regulating zcytor5 natural ligands or detecting cardiostrophin-1
 PT in blood
 PS Claim 1; Page 87-88; 55pp; English.
 CC The present sequence represents a zcytor5 variant protein. zcytor5
 CC is a cytokinin-like receptor. Soluble zcytor5 may be administered to
 CC down-regulate the effects of a growth and/or maintenance factor in
 CC thyroid, heart, and skeletal muscle for example to lessen the effect
 CC of cardiostrophin-1 on cardiac pathologies, so preventing heart
 CC enlargement. zcytor5 could be used to detect cardiostrophin-1 in the
 CC blood, and to discover other possible zcytor5 ligands. A probe
 CC comprising zcytor5 DNA or RNA can be used to determine the presence
 CC and integrity of the zcytor5 gene on chromosome 19. Antibodies and the
 CC anti-idiotypic antibody could be used to purify zcytor5 and
 CC therapeutically to modify zcytor5 ligand effects.
 SQ Sequence 389 AA;

Query Match 64.6%; Score 51; DB 39; Length 389;
 Best Local Similarity 60.0%; Pred. No. 6.97e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 226 lkdfifqaky 235
 QY 365 LKTPLFHSKY 374
 || ||:|

RESULT 14

ID W70853 standard; Protein: 389 AA.
 AC W70853;
 DT 17-MAR-1999 (first entry)
 DE Human zcytor5 variant.
 KW zcytor5; cytokinin-like receptor; down-regulation; growth factor;
 KW maintenance factor; thyroid; heart; skeletal muscle; cardiostrophin-1;
 KW cardiac pathology; heart enlargement; zcytor5 ligand; variant.
 OS Homo sapiens.
 PN W09849307-Al.
 PD 05-NOV-1998.
 PF 01-MAY-1998; U08865.
 PR 13-FEB-1998; US-074721.
 PR 01-MAY-1997; US-045287.
 PR 01-MAY-1997; US-850030.
 PR 13-FEB-1998; US-023890.
 PA (ZYMO) ZYMOGENETICS INC.
 PI Adams RL, Foster DC, Gilbert T, Jelmberg AC, Lehner JM,
 PI Lok S, Presnell SR, Whitmore TE;
 DR WPI: 99-034662/03.

PT New mammalian cytokinin-like receptor zcytor5 - useful for, e.g.
 PT down-regulating zcytor5 natural ligands or detecting cardiostrophin-1
 PT in blood
 PS Claim 1; Page 99-100; 55pp; English.
 CC The present sequence represents a zcytor5 variant protein. zcytor5
 CC is a cytokinin-like receptor. Soluble zcytor5 may be administered to
 CC down-regulate the effects of a growth and/or maintenance factor in
 CC thyroid, heart, and skeletal muscle for example to lessen the effect
 CC of cardiostrophin-1 on cardiac pathologies, so preventing heart
 CC enlargement. zcytor5 could be used to detect cardiostrophin-1 in the
 CC blood, and to discover other possible zcytor5 ligands. A probe
 CC comprising zcytor5 DNA or RNA can be used to determine the presence
 CC and integrity of the zcytor5 gene on chromosome 19. Antibodies and the
 CC anti-idiotypic antibody could be used to purify zcytor5 and
 CC therapeutically to modify zcytor5 ligand effects.
 SQ Sequence 389 AA;

Query Match 64.6%; Score 51; DB 39; Length 389;

Best Local Similarity 60.0%; Pred. No. 6.97e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 226 lkdfllfqak 235
|| ||:|
QY 365 LKTPLEFSKY 374

RESULT 15

ID W70846 standard; Protein; 389 AA.
AC W70846;
DT 17-MAR-1999 (first entry)
DE Human Zcytor5 variant.
KW Zcytor5; cytokinin-like receptor; down-regulation; growth factor;
KW maintenance factor; thyroid; heart; skeletal muscle; cardiostrophin-1;
KW cardiac pathology; heart enlargement; Zcytor5 ligand; variant.
OS Homo sapiens.
PN W09849307.A1.
PD 05-NOV-1998.
PF 01-MAY-1998; U08865.
PR 13-FEB-1998; US-074721.
PR 01-MAY-1997; US-045287.
PR 01-MAY-1997; US-850030.
PR 13-FEB-1998; US-023890.
PA (ZYMO) ZYMOGENETICS INC.
PI Adams RL, Foster DC, Gilbert T, Jelmsberg AC, Lehner JM,
PI Lok S, Presnell SR, Whitmore TE;
DR WPI: 99-034662/03.
PT New mammalian cytokinin-like receptor Zcytor5 - useful for, e.g.
PT down-regulating Zcytor5 natural ligands or detecting cardiostrophin-1
PT in blood
PS Claim 1; Page 89-90; 55pp; English.
CC The present sequence represents a Zcytor5 variant protein. Zcytor5
is a cytokinin-like receptor. Soluble Zcytor5 may be administered to
down-regulate the effects of a growth and/or maintenance factor in
thyroid, heart, and skeletal muscle for example to lessen the effect
of cardiostrophin-1 on cardiac pathologies, so preventing heart
enlargement. Zcytor5 could be used to detect cardiostrophin-1 in the
blood, and to discover other possible Zcytor5 ligands. A probe
comprising Zcytor5 DNA or RNA can be used to determine the presence
and integrity of the Zcytor5 gene on chromosome 19. Antibodies and the
anti-idiotypic antibody could be used to purify Zcytor5 and
therapeutically to modify Zcytor5 ligand effects.
SQ Sequence 389 AA;

Query Match 64.6%; Score 51; DB 39; Length 389;
Best Local Similarity 60.0%; Pred. No. 6.97e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 226 lkdfllfqak 235
|| ||:|
QY 365 LKTPLEFSKY 374

Search completed: Thu Jul 8 18:52:37 1999
Job time : 16 secs.

WQESRL (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:51:54 1999; MasPar time 4.99 Seconds
Tabular output not generated. 80.338 Million cell updates/sec

Title: >US-09-041-236-2
Description: (365-374) from US09041236.pap (22 of 45)
Perfect Score: 79
Sequence: 1 LKTPLEHSKY 10

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1.pir1 2.pir2 3.pir3 4.pir4

Statistics: Mean 24.310; Variance 31.620; scale 0.769

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	58	73.4	655	2	A40675 nuclear envelope prot	4.94e-01
2	56	70.9	209	2	T01336 hypothetical protein	1.26e-00
3	53	67.1	354	2	JU0457 peroxidase (EC 1.11.1	4.96e-00
4	53	67.1	481	2	B43674 protein kinase (EC 2.	4.96e-00
5	52	65.8	361	2	S23346 hypothetical protein	7.74e-00
6	52	65.8	364	2	I48188 gene Nkx6.1 protein	7.74e-00
7	52	65.8	516	2	S34525 hypothetical protein	7.74e-00
8	51	64.6	428	2	S47096 cytochrome-c oxidase	1.20e-01
9	51	64.6	509	2	S49349 cyprosin (EC 3.4.23.-	1.20e-01
10	51	64.6	722	2	S52205 mes1 protein - Leucon	1.20e-01
11	51	64.6	833	2	A47528 transcription factor	1.20e-01
12	51	64.6	1057	2	S45801 probable membrane pro	1.20e-01
13	51	64.6	1583	2	S59644 hypothetical protein	1.20e-01
14	50	63.3	335	2	A34290 membrane protein CBP3	1.85e-01
15	50	63.3	349	2	G64668 ADP-heptose-lps hept	1.85e-01
16	50	63.3	349	2	E71847 adp-heptose--lps hept	1.85e-01
17	50	63.3	352	2	T02875 ribosomal protein S4	1.85e-01
18	50	63.3	385	2	G70170 protein-glutamate met	1.85e-01
19	50	63.3	663	2	S74535 drug sensory protein	1.85e-01
20	49	62.0	80	1	OSRU7B cytochrome-c oxidase	2.84e-01
21	49	62.0	149	2	S48927 hypothetical protein	2.84e-01
22	49	62.0	223	2	F70469 enolase-phosphatase E	2.84e-01
23	49	62.0	288	2	F69973 transcription regulat	2.84e-01

24	49	62.0	418	2	S39833 hypothetical protein	2.84e-01
25	49	62.0	424	2	S09884 hypothetical protein	2.84e-01
26	49	62.0	448	2	JW0038 beta-glucosidase (EC	2.84e-01
27	49	62.0	510	2	I56242 lymphoid cell activat	2.84e-01
28	49	62.0	529	2	A47114 phosphoprotein phosph	2.84e-01
29	49	62.0	630	2	S48910 hypothetical protein	2.84e-01
30	49	62.0	1150	2	B47114 phosphoprotein phosph	2.84e-01
31	48	60.8	159	2	S40193 ribosomal protein L22	4.32e-01
32	48	60.8	259	1	DEECSF sorbitol-6-phosphate	4.32e-01
33	48	60.8	334	2	C22735 hypothetical nox2 pro	4.32e-01
34	48	60.8	351	2	B42009 FMLP-related receptor	4.32e-01
35	48	60.8	385	2	JN0535 integrase - Lactobaci	4.32e-01
36	48	60.8	413	2	JX0267 alpha-1-antiproteinas	4.32e-01
37	48	60.8	441	2	C59834 sensory transduction	4.32e-01
38	48	60.8	485	2	T03426 sucrose phosphorylase	4.32e-01
39	48	60.8	517	2	S21042 cytochrome-c oxidase	4.32e-01
40	48	60.8	529	2	F71547 probable oligopeptide	4.32e-01
41	48	60.8	551	2	S66695 probable membrane pro	4.32e-01
42	48	60.8	645	2	A44861 keratin, 67K type II	4.32e-01
43	48	60.8	664	2	F70483 glycyl-tRNA synthetas	4.32e-01
44	48	60.8	761	1	BBMS complement factor B p	4.32e-01
45	48	60.8	1049	1	S19421 ATP-dependent permeas	4.32e-01

ALIGNMENTS

RESULT 1
ENTRY nuclear envelope protein NDC1 - yeast (Saccharomyces cerevisiae)
TITLE protein YML031w
ALTERNATE_NAMES #formal_name Saccharomyces cerevisiae
ORGANISM 03-May-1994 #sequence_revision 03-May-1994 #text_change
DATE 01-May-1998
ACCESSIONS A40675; S49748; S42262
REFERENCE A40675
#authors Winey, M.; Hoyt, M.A.; Chan, C.; Goetsch, L.; Botstein, D.; Byers, B.
#journal J. Cell Biol. (1993) 122:743-751
#title NDC1: a nuclear periphery component required for yeast spindle pole body duplication.
#cross-references MUID:93352690
#accession A40675
#molecule_type DNA
#residues 1-655 #label WIN
#cross-references GB:X70281; NID:g4029; PID:g295764
REFERENCE S49741
#authors Badcock, K.; Churcher, C.
#submission submitted to the EMBL Data Library, November 1994
#accession S49748
#molecule_type DNA
#residues 1-655 #label BAD
#cross-references EMBL:Z46659; NID:g575680; PID:g575688; MIPS:YML031w
GENETICS
#gene SGD:NDC1
#cross-references SGD:S0004493; MIPS:YML031w
#map_position 13L
FUNCTION required for spindle pole body duplication
KEYWORDS nuclear membrane; transmembrane protein
FEATURE
34-34 #domain transmembrane #status predicted #label TM1\
59-79 #domain transmembrane #status predicted #label TM2\
90-110 #domain transmembrane #status predicted #label TM3\
117-137 #domain transmembrane #status predicted #label TM4\
191-211 #domain transmembrane #status predicted #label TM5\
225-245 #domain transmembrane #status predicted #label TM6\
514-533 #domain transmembrane #status predicted #label TM7\
SUMMARY #length 655 #molecular-weight 74133 #checksum 5334
Query Match 73.4%; Score 58; DB 2; Length 655;
Best Local Similarity 50.0%; Pred. No. 4.94e-01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```
Db 305 LRPIYHSKF 314
    |:::|::|:
Qy 365 LKTPLFHSKY 374

RESULT 2
ENTRY #type complete
TITLE hypothetical protein F6N15.9 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
        cress
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
ACCESSIONS T01336
REFERENCE #authors Ryan, E.; Edwards, J.; Pape, K.
#submission submitted to the EMBL Data Library, May 1998
#description The sequence of A. thaliana F6N15.
#accession T01336
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-209 #label RYA
#cross-references EMBL:AF069299; NID:g3193311; PID:g3193325
GENETICS
#map_position IV
#note F6N15.9
SUMMARY #length 209 #molecular-weight 23974 #checksum 1985

Query Match 70.9%; Score 56; DB 2; Length 209;
Best Local Similarity 50.0%; Pred. No. 1.26e+00;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 73 LKASLYRAKY 82
    |:::|::|:
Qy 365 LKTPLFHSKY 374

RESULT 3
ENTRY #type complete
TITLE peroxidase (EC 1.11.1.7) C - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
        cress
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
ACCESSIONS JU0457
REFERENCE #authors Incapruk, C.; Higashimura, N.; Yamamoto, K.; Okada, N.;
        Shimmyo, A.; Takano, M.
#journal Gene (1991) 98:237-241
#title Nucleotide sequences of two genomic DNAs encoding peroxidase
        of Arabidopsis thaliana.
#cross-references MUID:91200671
#accession JU0457
#molecule_type DNA
#residues 1-354 #label INT
#cross-references GB:M58380; NID:g166826; PID:g166827
GENETICS
#gene prxCa
#introns 78/3; 142/3; 199/1
CLASSIFICATION #superfamily peroxidase
KEYWORDS chromoprotein; heme; oxidoreductase
FEATURE
73.201 #binding_site heme iron (His) (axial ligands) #status
        predicted
SUMMARY #length 354 #molecular-weight 38941 #checksum 5820

Query Match 67.1%; Score 53; DB 2; Length 354;
Best Local Similarity 63.6%; Pred. No. 4.96e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 254 LRTPLVFNKY 264
    |::|::|:
Qy 365 LKTPLFHSKY 374

RESULT 4
ENTRY #type complete
TITLE protein kinase (EC 2.7.1.37) - human herpesvirus 2
ALTERNATE_NAMES #protein-serine/threonine kinase; US3 protein
        #formal_name human herpesvirus 2
ORGANISM #formal_name human herpesvirus 2
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
ACCESSIONS B43674
REFERENCE #authors McGeoch, D.J.; Moss, H.W.M.; McNab, D.; Frame, M.C.
        J. Gen. Virol. (1987) 68:19-38
#title DNA sequence and genetic content of the HindIII 1 region in
        the short unique component of the herpes simplex virus type
        2 genome: identification of the gene encoding glycoprotein
        G, and evolutionary comparisons.
#accession B43674
#molecule_type DNA
#residues 1-481 #label MCG
#cross-references EMBL:X04798; NID:g59900; PID:g59903
#experimental_source Strain HG52
CLASSIFICATION #superfamily kinase-related transforming protein; protein
        kinase homology
KEYWORDS ATP; phosphotransferase; serine/threonine-specific protein
        kinase; transforming protein
FEATURE
189-455 #domain protein kinase homology #label KIN\
197-205 #region protein kinase ATP-binding motif\
220 #active_site Lys #status Predicted
SUMMARY #length 481 #molecular-weight 52677 #checksum 6291

Query Match 67.1%; Score 53; DB 2; Length 481;
Best Local Similarity 66.7%; Pred. No. 4.96e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 473 LRLPLFHPK 481
    |::|::|:
Qy 365 LKTPLFHSK 373

RESULT 5
ENTRY #type complete
TITLE hypothetical protein 10.8 - Salmonella choleraesuis
ORGANISM #formal_name Salmonella choleraesuis
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
ACCESSIONS S23346
REFERENCE #authors Wang, L.; Romana, L.K.; Reeves, P.R.
        Genetics (1992) 130:429-443
#journal Molecular analysis of a Salmonella enterica group El rfb gene
        cluster: O antigen and the genetic basis of the major
        polymorphism.
#cross-references MUID:92201631
#accession S23346
#status preliminary
#molecule_type DNA
#residues 1-361 #label WAN
#cross-references EMBL:X0665; NID:g46984; PID:g46989
#note the source is designated as Salmonella enterica
SUMMARY #length 361 #molecular-weight 40673 #checksum 8156

Query Match 65.8%; Score 52; DB 2; Length 361;
Best Local Similarity 60.0%; Pred. No. 7.74e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 78 LKGFLENKY 87
    |::|::|:
Qy 365 LKTPLFHSKY 374
```

```

6
RESULT 6
ENTRY I48188 #type complete
TITLE gene NKx5.1 protein - golden hamster
ORGANISM #formal_name Mesocricetus auratus #common_name golden hamster
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
17-Oct-1997
ACCESSIONS I48188
REFERENCE I48185
#authors Rudnick, A.; Ling, T.Y.; Odagiri, H.; Rutter, W.J.; German,
M.S.
#journal Proc. Natl. Acad. Sci. U.S.A. (1994) 91:12203-12207
#title Pancreatic beta cells express a diverse set of homeobox
genes.
#cross-references MUID:95083670
#accession I48188
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-364 #label RES
#cross-references EMBL:X81409; NID:g587466; PID:g587467
GENETICS
#gene NKx5.1
#classification #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS DNA binding; homeobox; nucleus; transcription regulation
FEATURE
237-293
SUMMARY #domain homeobox homology #label HOX #length 364 #molecular-weight 37646 #checksum 6437

Query Match 65.8%; Score 52; DB 2; Length 364;
Best Local Similarity 50.0%; Pred. No. 7.74e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 31 MKTPLYPAY 40
QY 365 LKTPLEHFSKY 374

7
RESULT 7
ENTRY S34525 #type complete
TITLE hypothetical protein 516 (rps3 3' region) - Euglena gracilis
chloroplast
ORGANISM #formal_name chloroplast Euglena gracilis
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
24-Sep-1998
ACCESSIONS S34892; S26088; S34525
REFERENCE S34862
#authors Hallick, R.B.; Hong, L.; Drager, R.G.; Favreau, M.R.;
Monfort, A.; Orsat, B.; Spielmann, A.; Stutz, E.
#journal Nucleic Acids Res. (1993) 21:3537-3544
#title Complete sequence of Euglena gracilis chloroplast DNA.
#cross-references MUID:93347989
#accession S34892
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-516 #label HAL2
#cross-references EMBL:X70810; NID:g415327; PID:e222140; PID:g1105109
#experimental_source strain Pringsheim Z
#note the nucleotide sequence was submitted to the EMBL Data
Library, January 1993
REFERENCE S26086
#authors Hallick, R.B.
#submission submitted to the EMBL Data Library, March 1992
#accession S26088
#molecule_type DNA
#residues 1-516 #label HAL
#cross-references EMBL:Z11874; NID:g14353; PID:e223926; PID:g1208513
GENETICS
#genome chloroplast
#start_codon TTG
#introns 17/2; 18/2; 22/3; 228/3
#classification #superfamily Euglena gracilis chloroplast hypothetical
protein 516
KEYWORDS chloroplast
SUMMARY #length 516 #molecular-weight 64333 #checksum 5374

Query Match 65.8%; Score 52; DB 2; Length 516;
Best Local Similarity 60.0%; Pred. No. 7.74e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 412 LKSHLYSSKY 421
QY 365 LKTPLEHFSKY 374

8
RESULT 8
ENTRY S47096 #type complete
TITLE cytarase (EC 3.4.23.-) - cardoon
ORGANISM #formal_name Cynara cardunculus #common_name cardoon
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
29-Aug-1997
ACCESSIONS S47096
REFERENCE S47096
#authors Cordeiro, M.C.; Xue, Z.T.; Pietrzak, M.; Pais, M.S.;
Brodelius, P.
#submission submitted to the EMBL Data Library, November 1992
#description Tissue specific expression of cynarase (s) genes in flowers
of Cynara cardunculus.
#accession S47096
#status preliminary
#molecule_type mRNA
#residues 1-428 #label COR
#cross-references EMBL:X69193
COMMENT The pair of saposin repeat homology domains tagged SAPI and SAP2
represent a cyclical permutation of a single saposin repeat.
CLASSIFICATION #superfamily oryzasin; saposin repeat homology
KEYWORDS aspartic proteinase; hydrolase
FEATURE
236-281
SUMMARY #domain saposin repeat homology #status atypical #label
SAP1\
#domain saposin repeat homology #status atypical #label
SAP2\
#active_site Asp #status predicted
#length 428 #molecular-weight 46469 #checksum 946

Query Match 64.6%; Score 51; DB 2; Length 428;
Best Local Similarity 100.0%; Pred. No. 1.20e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 43 LFHSKY 48
QY 369 LFHSKY 374

9
RESULT 9
ENTRY S49349 #type complete
TITLE cyprosin (EC 3.4.23.-) - cardoon
ORGANISM #formal_name Cynara cardunculus #common_name cardoon
DATE 16-Feb-1995 #sequence_revision 12-May-1995 #text_change
08-Sep-1997
ACCESSIONS S49349
REFERENCE S49349
#authors Pietrzak, M.; Brodelius, P.; Pais, M.S.
#submission submitted to the EMBL Data Library, September 1994
#accession S49349
#status preliminary
#molecule_type mRNA
#residues 1-509 #label PIE
#cross-references EMBL:X81984; NID:g556818; PID:g556819
COMMENT The pair of saposin repeat homology domains tagged SAPI and SAP2
represent a cyclical permutation of a single saposin repeat.
CLASSIFICATION #superfamily oryzasin; saposin repeat homology
KEYWORDS aspartic proteinase; hydrolase
FEATURE
316-361
SUMMARY #domain saposin repeat homology #status atypical #label
SAP1\
#domain saposin repeat homology #status atypical #label
SAP2\
#length 420
SUMMARY #domain saposin repeat homology #status atypical #label
SAP2\
#length 420
```

```

103,290      #active_site Asp #status predicted
SUMMARY      #length 509 #molecular-weight 55176 #checksum 7843

Query Match      64.6%; Score 51; DB 2; Length 509;
Best Local Similarity 100.0%; Pred. No. 1.20e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 123 LFHSKY 128
|||||
Oy 369 LFHSKY 374

RESULT 10
ENTRY      #type complete
TITLE      mes1 protein - Leuconostoc mesenteroides
ORGANISM   #formal_name Leuconostoc mesenteroides
DATE       07-May-1995 #sequence_revision 21-Jul-1995 #text_change
17-Jul-1998
ACCESSIONS S52205
REFERENCE   S52204
#authors   Fremaux, C.; Hechard, A.; Sepulchre, A.; Cenatiempo, Y.
#submission submitted to the EMBL Data Library, September 1994
#description Organization of mesentericin Y105 genetic determinants.
#accession S52205
#status    preliminary
#molecule_type DNA
#residues  1-722 #label FRE
#cross-references EMBL:X81803; NID:g609082; PID:g609084
CLASSIFICATION #superfamily unassigned ATP-binding cassette proteins;
ATP-binding cassette homology
KEYWORDS    ATP; P-loop
FEATURE     #domain ATP-binding cassette homology #label ABC\
505-698      #region nucleotide-binding motif A (P-loop)
522-529      #length 722 #molecular-weight 79915 #checksum 8276
SUMMARY

Query Match      64.6%; Score 51; DB 2; Length 722;
Best Local Similarity 66.7%; Pred. No. 1.20e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 2 VKTPMFHKK 10
|||||
Oy 365 LKTPLFHKK 373

RESULT 11
ENTRY      #type complete
TITLE      transcription factor MBP1 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein D2362; protein YDL056w
ORGANISM   #formal_name Saccharomyces cerevisiae
DATE       03-May-1994 #sequence_revision 03-May-1994 #text_change
17-Mar-1999
ACCESSIONS A47528; B47528; S59375; S67591; S62745; S37404
REFERENCE   A47528
#authors   Koch, C.; Moll, T.; Neuberg, M.; Ahorn, H.; Nasmyth, K.
#journal   Science (1993) 261:1551-1557
#title     A role for the transcription factors Mbp1 and Swi4 in
progression from G1 to S phase.
#cross-references MUID:93383264
#accession A47528
#status    nucleic acid sequence not shown
#molecule_type DNA
#residues  1-833 #label KOC
#cross-references GB:X74158; NID:g402792; PID:g402793
#accession B47528
#molecule_type protein
#residues  'X', 372-377, 'X', 379-381, 'X', 383-387 #label K02
REFERENCE   S59374
#authors   Benton, B.K.; Plump, S.D.; Roos, J.; Lennarz, W.J.; Cross,
F.R.
#submission submitted to the EMBL Data Library, January 1995
#description Overexpression of S. cerevisiae G1 cyclins restores viability
of aigl N-glycosylation mutants.
#accession S59375
#molecule_type DNA
#residues  460-833 #label BEN
#cross-references EMBL:U19608; NID:g1292897; PID:g639493
REFERENCE   S67587
#authors   Bloeker, H.; Brandt, P.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67591
#molecule_type DNA
#residues  1-833 #label BLO
#cross-references EMBL:Y74104; NID:g1431054; PID:e253002; PID:g1431055;
MIPS:YDL056w
#experimental_source strain S288C
REFERENCE   S62744
#authors   Benton, B.K.; Plump, S.D.; Roos, J.; Lennarz, W.J.; Cross,
F.R.
#journal   Curr. Genet. (1996) 29:106-113
#title     Over-expression of S. cerevisiae G(1) cyclins restores the
viability of aigl N-glycosylation mutants.
#cross-references MUID:96418871
#accession S62745
#status    nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues  460-833 #label BEW
#cross-references EMBL:U19608; NID:g1292897; PID:g639493
#note      the nucleotide sequence was submitted to the EMBL Data
Library, January 1995
GENETICS
#gene       SGD:MBP1
#cross-references SGD:S0002214; MIPS:YDL056w
#map_position 4L
CLASSIFICATION #superfamily unassigned ankyrin repeat proteins; ankyrin
repeat homology
KEYWORDS    DNA binding; nucleus; transcription regulation
FEATURE     #domain ankyrin repeat homology #label AN1
394-426      #length 833 #molecular-weight 93907 #checksum 2204
SUMMARY

Query Match      64.6%; Score 51; DB 2; Length 833;
Best Local Similarity 40.0%; Pred. No. 1.20e+01;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 433 MRSLFHNSY 442
|||||
Oy 365 LKTPLFHNSY 374

RESULT 12
ENTRY      #type complete
TITLE      probable membrane protein YBL066c - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES hypothetical protein YBL0501; hypothetical protein YBL0526
ORGANISM   #formal_name Saccharomyces cerevisiae
DATE       26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change
17-Mar-1999
ACCESSIONS S45801; S45802; S39824; S37325
REFERENCE   S45782
#authors   Dubois, E.; El Bakkoury, M.; Glansdorff, N.; Messenguy, F.;
Pierard, A.; Scherens, B.; Vicerondeils, F.
#submission submitted to the Protein Sequence Database, August 1994
#accession S45801
#molecule_type DNA
#residues  1-788 #label DUB
#cross-references EMBL:Z35827; MIPS:YBL066c
#experimental_source strain S288C
REFERENCE   S45802
#authors   Contreras, R.; Fiers, W.; Logghe, M.; Molemans, F.
#submission submitted to the Protein Sequence Database, August 1994
#accession S45802
#molecule_type DNA
#residues  285-1057 #label CON
#cross-references EMBL:Z35827; MIPS:YBL066c
#experimental_source strain S288C

```



```

REFERENCE          S39824
#authors           Scherens, B.; el Bakkoury, M.; Vierendeels, F.; Dubois, E.;
#journal           Messenguy, F. 1355-1371
#title            Yeast (1993) 9:1355-1371
#cross-references  Sequencing and functional analysis of a 32 560 bp segment on
                    the left arm of yeast chromosome II. Identification of 26
                    open reading frames, including the KIP1 and SEC17 genes.
#accession         MUD:94205266
#molecule_type   DNA
#residues          1-788 #label SCH
#cross-references EMBL:23261
#experimental_source strain S288C

GENETICS
#gene             SGD:SEF1
#map_position     2L
#cross-references SGD:S000162; MIPS:YBL066c
CLASSIFICATION    #superfamily unassigned GAL4-type zinc cluster proteins; GAL4
                    zinc binuclear cluster homology
KEYWORDS          DNA binding; nucleus; transmembrane protein; zinc finger
FEATURE           52-92
#domain GAL4 zinc binuclear cluster homology #label
                    GAL4\
#region zinc finger\
#domain transmembrane #status predicted #label TM1\
#domain transmembrane #status predicted #label TM2\
#domain transmembrane #status predicted #label TM3\
#length 1057 #molecular-weight 118191 #checksum 7973
SUMMARY
Query Match      64.6%; Score 51; DB 2; Length 1057;
Best Local Similarity 66.7%; Pred. No. 1.20e+01;
Matches          6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1038 RTPLFSIY 1046
   :||||:|
Qy 366 KTPLFHSKY 374

RESULT 13
ENTRY   S59644 #type complete
TITLE   hypothetical protein SPAC31A2.05c - fission yeast
        (Schizosaccharomyces pombe)
ORGANISM #formal name Schizosaccharomyces pombe
DATE     14-Jan-1996 #sequence_revision 19-Apr-1996 #text_change
        31-Oct-1997
ACCESSIONS S59644
REFERENCE  S58093
#authors   Devlin, K.; Churcher, C.M.
#submission submitted to the EMBL Data Library, July 1995
#accession S59644
#status    preliminary
#molecule_type DNA
#residues  1-1583 #label DEV
#cross-references EMBL:250113; NID:g914878; PID:g914883
GENETICS
#introns  33/1; 98/2; 543/3; 699/3; 1294/2; 1339/3; 1558/3
SUMMARY   #length 1583 #molecular-weight 180202 #checksum 4709

Query Match      64.6%; Score 51; DB 2; Length 1583;
Best Local Similarity 54.5%; Pred. No. 1.20e+01;
Matches          6; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 446 LKAPLFOTLOY 456
   :||||:|
Qy 365 KTPLFHS-KY 374

RESULT 14
ENTRY   A34290 #type complete
TITLE   membrane protein CBP3 - Yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein P1775; protein YPL215w
ORGANISM #formal name Saccharomyces cerevisiae
DATE     22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change

```

```

06-Feb-1998
A34290; S55234
A34290
#authors          Wu, M.; Tzagoloff, A.
#journal          J. Biol. Chem. (1989) 264:11122-11130
#title            Identification and characterization of a new gene (CBP3)
                    required for the expression of yeast coenzyme
                    QH-2-cytochrome c reductase.
#cross-references MUD:89291856
#accession        A34290
#molecule_type   DNA
#residues          1-335 #label WUM
#cross-references GB:J04830; NID:g171170; PID:g171171
REFERENCE S65202
#authors          Rieger, M.; Mueller-Auer, S.; Schaefer, M.
#submission       submitted to the Protein Sequence Database, May 1996
#accession        S65234
#molecule_type   DNA
#residues          1-335 #label RIE
#cross-references EMBL:273571; NID:g1370445; PID:e247067; PID:g1370446;
                    MIPS:YPL215w
#experimental_source strain S288C (AB972)

GENETICS
#gene             SGD:CBP3
#cross-references SGD:S0006136; MIPS:YPL215w
#map_position     16L
#genome          nuclear
KEYWORDS          membrane protein; mitochondrial
SUMMARY           #length 335 #molecular-weight 39083 #checksum 2407

Query Match      63.3%; Score 50; DB 2; Length 335;
Best Local Similarity 77.8%; Pred. No. 1.85e+01;
Matches          7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 73 KIPLAHSKY 81
   :||||:|
Qy 366 KTPLFHSKY 374

RESULT 15
ENTRY   G64668 #type complete
TITLE   ADP-heptose-1ns heptosyltransferase II - Helicobacter pylori
        (strain 26595)
ORGANISM #formal_name Helicobacter pylori
DATE     09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
        10-Oct-1997
ACCESSIONS G64668
REFERENCE  A64520
#authors    Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
            Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
            H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
            J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
            Richardson, R.; Dodson, R.; Khalak, H.G.; Glodek, A.;
            McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
            Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
            Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
            Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
            W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
            C.M.; Venter, J.C.
#journal      Nature (1997) 388:539-547
#title        The complete genome sequence of the gastric pathogen
            Helicobacter pylori.
#cross-references MUD:97394467
#accession     G64668
#status        preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues      1-349 #label TOM
#cross-references GB:AE000625; GB:AE000511; NID:g2314349; PID:g2314352;
            TIGR:HP1191
SUMMARY        #length 349 #molecular-weight 39447 #checksum 157

Query Match      63.3%; Score 50; DB 2; Length 349;

```

Best Local Similarity 63.6%; Pred. No. 1.85e+01;
Matches 7; Conservative 1; Mismatches 2; Indels 1; Gaps 1;
Db 233 LKNPSLFHAY 243
||| ||| |
QY 365 LKTP-LFHSKY 374

Search completed: Thu Jul 8 18:52:03 1999
Job time : 9 secs.

MPERCH_PP

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:50:18 1999; MasPar time 3.48 Seconds

Tabular output not generated. 81.116 Million cell updates/sec

Title: >US-09-041-236-2
Description: (365-374) from US09041236.pep (22 of 45)
Perfect Score: 79
Sequence: 1 LKTPLEFSKY 10

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 24.968; Variance 27.987; scale 0.892

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	58	73.4	655	1	NDCL1_YEAST NUCLEAR ENVELOPE PROTEIN	1.07e+01
2	53	67.1	354	1	PERC_ARATH PERC ARATH NUCLEAR ENVELOPE PROTEIN	1.44e+00
3	53	67.1	481	1	KRL_HSV2H SERINE/THREONINE-PROTEIN	1.44e+00
4	52	65.8	293	1	MOTY_VIBPA SODIUM-TYPE FLAGELLAR	2.38e+00
5	52	65.8	516	1	YCX4_EUGR HYPOTHETICAL 64.3 KD P	2.38e+00
6	51	64.6	473	1	CYPI_CYNCA CYPROSIN PRECURSOR (EC	3.90e+00
7	51	64.6	722	1	MESD_LEUME MESENTERICIN Y105 TRAN	3.90e+00
8	51	64.6	833	1	MBP1_YEAST TRANSCRIPTION FACTOR M	3.90e+00
9	51	64.6	1057	1	SEF1_YEAST SUPPRESSOR PROTEIN SEF	3.90e+00
10	51	64.6	1583	1	Y445_SCHPO HYPOTHETICAL 180.2 KD	3.90e+00
11	50	63.3	335	1	CBP3_YEAST CBP3 PROTEIN PRECURSOR	6.34e+00
12	50	63.3	663	1	DSBP_SYNY3 DRUG SENSORY PROTEIN A	6.34e+00
13	49	62.0	80	1	COXM_HUMAN CYTOCHROME C OXIDASE P	1.02e+01
14	49	62.0	149	1	YHE1_YEAST HYPOTHETICAL 17.6 KD P	1.02e+01
15	49	62.0	418	1	YBF5_YEAST HYPOTHETICAL 47.4 KD P	1.02e+01
16	49	62.0	424	1	ULB7_HCMVA HYPOTHETICAL PROTEIN U	1.02e+01
17	49	62.0	448	1	BGLB_PAEPO BETA-GLUCOSIDASE B (EC	1.02e+01
18	49	62.0	477	1	Y142_NPVAC HYPOTHETICAL 55.4 KD P	1.02e+01
19	49	62.0	510	1	CD39_HUMAN VASCULAR ATP-DIPHOSPHO	1.02e+01
20	49	62.0	539	1	2ACB_HUMAN PROTEIN PHOSPHATASE PP	1.02e+01
21	49	62.0	520	1	YHX1_YEAST HYPOTHETICAL 71.4 KD P	1.02e+01
22	49	62.0	1150	1	2ACA_HUMAN PROTEIN PHOSPHATASE PP	1.02e+01
23	48	60.8	80	1	COXM_MOUSE CYTOCHROME C OXIDASE P	1.64e+01

24	48	60.8	159	1	RL22_THEMA 50S RIBOSOMAL PROTEIN	1.64e+01
25	48	60.8	259	1	SRLD_ECOLI SORBITOL-6-PHOSPHATE 2	1.64e+01
26	48	60.8	348	1	FML1_PONPY FMLP-RELATED RECEPTOR	1.64e+01
27	48	60.8	348	1	FML1_PANTR FMLP-RELATED RECEPTOR	1.64e+01
28	48	60.8	348	1	FML1_GORGO FMLP-RELATED RECEPTOR	1.64e+01
29	48	60.8	351	1	FML1_HUMAN FMLP-RELATED RECEPTOR	1.64e+01
30	48	60.8	517	1	QOX1_SULAC QUINOL OXIDASE POLYPEP	1.64e+01
31	48	60.8	645	1	K2E2_HUMAN KERATIN, TYPE II CYTOS	1.64e+01
32	48	60.8	761	1	CFAB_MOUSE COMPLEMENT FACTOR B PR	1.64e+01
33	48	60.8	1049	1	ADP1_YEAST PROBABLE ATP-DEPENDENT	1.64e+01
34	47	59.5	129	1	VAL2_PYMV AL2 PROTEIN.	2.60e+01
35	47	59.5	159	1	HS21_SOYBN 17.9 KD CLASS II HEAT	2.60e+01
36	47	59.5	347	1	PER2_ARMRU PEROXIDASE C2 PRECURSO	2.60e+01
37	47	59.5	353	1	PERA_ARMRU PEROXIDASE CIA PRECURS	2.60e+01
38	47	59.5	412	1	CCA_SULSH TRNA NUCLEOTIDYLTRANSFER	2.60e+01
39	47	59.5	441	1	DCDA_BACSU DIAMINOPIMELATE DECARB	2.60e+01
40	47	59.5	447	1	YK5_YEAST HYPOTHETICAL 50.3 KD P	2.60e+01
41	47	59.5	592	1	INV2_DAUCA BETA-FRUCTOFURANOSIDAS	2.60e+01
42	47	59.5	656	1	YC26_PORPU HYPOTHETICAL SENSOR-LI	2.60e+01
43	47	59.5	664	1	ER72_CAEEL PROBABLE ERP72 PROTEIN	2.60e+01
44	47	59.5	921	1	YIB7_YEAST HYPOTHETICAL 105.5 KD	2.60e+01
45	47	59.5	1137	1	MSB1_YEAST MORPHOGENESIS-RELATED	2.60e+01

ALIGNMENTS

RESULT 1
ID NDC1_YEAST STANDARD; PRT; 655 AA.
AC P32500;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE NUCLEAR ENVELOPE PROTEIN NDC1.
GN NDC1 OR YML031W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE; 93352690.
RA WINEY M., HOYT M.A., CHAN C., GOETSCH L., BOTSTEIN D., BYERS B.:
RT "NDC1: a nuclear periphery component required for yeast spindle pole
body duplication.";
RL J. CELL BIOL. 122:743-751(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA BADCOCK K., CHURCHER C., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- FUNCTION: ESSENTIAL FOR MITOTIC VIABILITY. REQUIRED FOR SPINDLE
POLE BODY (SPB) DUPLICATION IN G1. MAY INSERT THE NASCENT SPB INTO
THE NUCLEAR ENVELOPE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. NUCLEAR ENVELOPE.
CC LOCALIZATION TO THE ER IS ALSO POSSIBLE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X70281; G295764; -
DR EMBL; 246659; G575686; -
DR PIR; A40675; A40675.
DR SGD; L0001237; NDC1.
KW NUCLEAR PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.

```
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 225 245 POTENTIAL.
FT TRANSMEM 514 533 POTENTIAL.
SQ SEQUENCE 655 AA; 74133 MW; 2213B806 CRC32;
Query Match 73.4%; Score 58; DB 1; Length 655;
Best Local Similarity 50.0%; Pred. No. 1.07e-01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 305 LRPIYHSKF 314
|:::|::|:
QY 365 LKTPLFHSKY 374

RESULT 2
ID PERC_ARATH STANDARD; PRT; 354 AA.
AC P24101;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE NEUTRAL PEROXIDASE C PRECURSOR (EC 1.11.1.7).
GN PRXA.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 91200671.
RX TAKANO M.;
RA "Nucleotide sequences of two genomic DNAs encoding peroxidase of
Arabidopsis thaliana.";
RL GENE 98:237-241(1991).
CC -!- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,
CC BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENCE RESPONSE TOWARD
CC WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE
CC DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.
CC -!- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.
CC -!- COFACTOR: THIS PROTEIN BINDS HEME.
CC -!- TISSUE SPECIFICITY: ROOTS.
CC -!- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M58380; G166827; -
DR PIR; JU0437; JU0457.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR PFAM; PF00141; peroxidase; 1.
DR HSP; P00433; 2ATJ.
KW OXIDOREDUCTASE; GLYCOPROTEIN; PEROXIDASE; HEME; MULTIGENE FAMILY;
KW SIGNAL.
FT SIGNAL 1 21
FT CHAIN 22 354 NEUTRAL PEROXIDASE C.
FT ACT_SITE 69 69 BY SIMILARITY.
FT ACT_SITE 73 73 DISTAL HISTIDINE (BY SIMILARITY).
FT ACT_SITE 201 201 PROXIMAL HISTIDINE (HEME AXIAL LIGAND).
FT ACT_SITE 42 122 BY SIMILARITY.
FT DISULFID 75 80 BY SIMILARITY.
FT DISULFID 128 332 BY SIMILARITY.
FT DISULFID 208 240 BY SIMILARITY.
FT CARBOHYD 88 88 POTENTIAL.
FT CARBOHYD 217 217 POTENTIAL.
FT CARBOHYD 229 229 POTENTIAL.
FT CARBOHYD 245 245 POTENTIAL.
FT CARBOHYD 286 286 POTENTIAL.

SQ SEQUENCE 354 AA; 38941 MW; 24986778 CRC32;
Query Match 67.1%; Score 53; DB 1; Length 354;
Best Local Similarity 63.6%; Pred. No. 1.44e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 254 LRTPLVFNDKY 264
|::|:|::|:
QY 365 LKTPLFHSKY 374

RESULT 3
ID KRL_HSV2H STANDARD; PRT; 481 AA.
AC P13287;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SERINE/THREONINE-PROTEIN KINASE (EC 2.7.1.-).
GN US3.
OS HERPES SIMPLEX VIRUS (TYPE 2 / STRAIN HG52).
OC VIRUSES; DSDNA VIRUSES; NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; SIMPLEXVIRUS.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 87111457.
RX MCGEOCH D.J., MOSS H.W.M., MCNAB D., FRAME M.C.;
RA "DNA sequence and genetic content of the HindIII 1 region in the
RT short unique component of the herpes simplex virus type 2 genome:
RT identification of the gene encoding glycoprotein G, and evolutionary
RT comparisons.";
RL J. GEN. VIROL. 68:19-38(1987).
[2]
RN SEQUENCE FROM N.A.
RP DOLAN A.;
RA SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
[3]
RN REVIEW.
RX MEDLINE: 89369695.
RA LEADER D.P., PURVES F.C.;
RT "The herpesvirus protein kinase: a new departure in protein
RT phosphorylation?";
RL TRENDS BIOCHEM. SCI. 13:244-246(1988).
CC -!- FUNCTION: PHOSPHORYLATES UL34 PROTEIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X04798; G59903; -
DR EMBL; 286099; E304196; -
DR PIR; B43674; B43674.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinaase; 1.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING.
FT DOMAIN 191 478 PROTEIN KINASE.
FT NP_BIND 197 205 ATP (BY SIMILARITY).
FT BINDING 220 220 ATP (BY SIMILARITY).
FT ACT_SITE 305 305 BY SIMILARITY.
SQ SEQUENCE 481 AA; 52677 MW; 558D65C CRC32;
Query Match 67.1%; Score 53; DB 1; Length 481;
Best Local Similarity 66.7%; Pred. No. 1.44e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 473 LRLPLFHPK 481
|:|::|:|
QY 365 LKTPLFHSK 373
```

RESULT 4
ID MOTY_VIBPA STANDARD; PRT; 293 AA.
AC P46233;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SODIUM-TYPE FLAGELLAR PROTEIN MOTY PRECURSOR.
GN MOTY.
OS VIBRIO PARAHAEEMOLYTICUS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIONACEAE; VIBRIO.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB22;
RX MEDLINE; 94292449.
RA MCCARTER L.L.;
RT "MotY, a component of the sodium-type flagellar motor.";
RL J. BACTERIOL. 175:4219-4225(1994).
CC -!- FUNCTION: PLAYS THE ROLE OF A STATOR IN THE SODIUM FLAGELLAR
CC MOTOR, STABILIZING THE FORCE-GENERATING UNIT THROUGH DIRECT
CC INTERACTION WITH THE CELL WALL.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U06949; G497129; -;
DR PROSITE; PS01068; OMPA; 1.
DR PFAM; PF00691; OmpA; 1.
KW FLAGELLA; SIGNAL.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 293 SODIUM-TYPE FLAGELLAR PROTEIN MOTY.
FT DOMAIN 220 264 OMPA-LIKE.
FT SEQUENCE 293 AA; 33406 MW; C95EB203 CRC32;
Query Match 65.8%; Score 52; DB 1; Length 293;
Best Local Similarity 60.0%; Pred. No. 2.38e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 153 LSSVLFQSKY 162
| : |||||
Qy 365 LKTPLFHSKY 374
RESULT 5
ID YC4_EUGGR STANDARD; PRT; 516 AA.
AC P30397;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 64.3 KD PROTEIN IN RPS3 3'REGION (ORF516).
OS EUGLENA GRACILIS.
OC CHLOROPLAST.
CC EUKARYOTA; EUGLENOZOEA; EUGLENIDA; EUGLENALES; EUGLENA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z;
RX MEDLINE; 93347989.
RA HALLICK R.B., HONG L., DRAGER R.G., FAVREAU M.R., MONFORT A.,
RA ORSAT B., SPIELMANN A., STUTZ E.;
RT "Complete sequence of Euglena gracilis chloroplast DNA.";
RL NUCLEIC ACIDS RES. 21:3537-3544(1993).
CC -!- SIMILARITY: TO GROUP II INTRON MATURASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X70810; E1204055; -;
DR EMBL; Z11874; E1204062; -;
DR PIR; S26088; S26088.
DR PIR; S34525; S34525.
KW CHLOROPLAST; HYPOTHETICAL PROTEIN.
SQ SEQUENCE 516 AA; 64333 MW; CE19DE9F CRC32;
Query Match 65.8%; Score 52; DB 1; Length 516;
Best Local Similarity 60.0%; Pred. No. 2.38e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 412 LKSHLYSSKY 421
| : |||||
Qy 365 LKTPLFHSKY 374
RESULT 6
ID CYP1_CYNCA STANDARD; PRT; 473 AA.
AC P40782;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE CYPROSIN PRECURSOR (EC 3.4.23.-) (FRAGMENT).
GN CYPROI.
OS CYNARA CARDUNCULUS (CARDUON).
CC EUKARYOTA; VIRIDIPALTA; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
CC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
CC ASTERIDAE; ASTERALES; ASTERACEAE; CYNARA.
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 178-186.
RC TISSUE=FLOWER BUDS;
RX MEDLINE; 94250836.
RA CORDEIRO M.C., XUE Z.-T., PIETRZAK M., PAIS M.S., BRODELIUS P.E.;
RT "Isolation and characterization of a cDNA from flowers of Cynara
RT cardunculus encoding cyprosin (an aspartic proteinase) and its use to
RT study the organ-specific expression of cyprosin.";
RL PLANT MOL. BIOL. 24:733-741(1994).
CC -!- TISSUE SPECIFICITY: MOSTLY PRESENT IN THE VIOLET PARTS OF STYLES
CC AND COROLLAS OF MATURE FLOWERS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN EARLY STAGES OF FLORAL
CC DEVELOPMENT AND SWITCHED OFF AT MATURATION OF THE FLOWER.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X69193; G509163; ALT SEQ.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR PFAM; PF00026; asp; 1.
DR HSSP; P00797; ZREN.
KW HYDROLASE; ASPARTYL PROTEASE; ZYMAGEN; GLYCOPROTEIN.
RN NON_TER 1 1
FT PROPEP <1 33 POTENTIAL.
FT CHAIN 34 473 BY SIMILARITY.
FT ACT_SITE 69 69 BY SIMILARITY.
FT ACT_SITE 256 256 BY SIMILARITY.
FT DISULFID 82 88 POTENTIAL.
FT DISULFID 247 251 POTENTIAL.
FT DISULFID 392 429 POTENTIAL.
FT CARBOHYD 364 364 POTENTIAL.
FT CARBOHYD 411 411 POTENTIAL.

SQ SEQUENCE 473 AA; 51564 MW; 563EB410 CRC32;
Query Match 64.6%; Score 51; DB 1; Length 473;
Best Local Similarity 100.0%; Pred. No. 3.90e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 89 LFHSKY 94
| | | | |
QY 369 LFHSKY 374

RESULT 7
ID MSRD_LEUME STANDARD; PRT; 722 AA.
AC Q10418;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE MESENTERICIN Y105 TRANSPORT ATP-BINDING PROTEIN MSRD.
GN MSRD.
OS LEUCONOSTOC MENSENTEROIDES.
OG PLASMID PHY30.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; LACTOBACILLACEAE;
OC LEUCONOSTOC.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Y105;
RX MEDLINE; 96004463.
RA FREMAUX C., HECHARD A., CENATIEMPO Y.;
RT "Mesentericin Y105 gene clusters in Leuconostoc mesenteroides Y105";
RL MICROBIOLOGY 141:1637-1645(1995).
CC -1- FUNCTION: INVOLVED IN THE EXPORT PROCESS OF THE BACTERIOCIN
CC MESENTERICIN Y105.
CC SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). BELONGS TO THE HLYB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X81803; G609084; -
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PFAM; PF00005; ABC_tran; 1.
DR PFAM; PF00684; ABC_membrane; 1.
DR HSP; P13569; INBD.
DR PLASMID; TRANSMEMBRANE; TRANSPORT; ATP-BINDING.
FT TRANSMEM 171 191 POTENTIAL.
FT TRANSEM 210 230 POTENTIAL.
FT TRANSEM 242 262 POTENTIAL.
FT TRANSEM 287 307 POTENTIAL.
FT TRANSEM 311 331 POTENTIAL.
FT TRANSEM 401 421 POTENTIAL.
FT TRANSEM 429 449 POTENTIAL.
FT TRANSEM 518 538 POTENTIAL.
FT NP_BIND 522 529 ATP (POTENTIAL).
SQ SEQUENCE 722 AA; 79916 MW; A327FD53 CRC32;

Query Match 64.6%; Score 51; DB 1; Length 722;
Best Local Similarity 66.7%; Pred. No. 3.90e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 2 VKTDMFHK 10
| | | | |
QY 365 LKTPLFHFK 373

RESULT 8
ID MBP1_YEAST STANDARD; PRT; 833 AA.
AC P39678;

SQ SEQUENCE 473 AA; 51564 MW; 563EB410 CRC32;
Query Match 64.6%; Score 51; DB 1; Length 833;
Best Local Similarity 40.0%; Pred. No. 3.90e+00;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 433 MRSSLFHSKY 442
| | | | |
QY 365 LKTPLFHFK 374

DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION FACTOR MBP1 (MBF SUBUNIT P120).
GN MBP1 OR YDL056W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 372-387.
RC STRAIN-K1107;
RX MEDLINE; 93383264.
RA KOCH C., MOLL T., NEUBERG M., AHORN H., NASMYTH K.;
RT "A role for the transcription factors Mbpl and Swi4 in progression
RT from G1 to S phase";
RL SCIENCE 261:1551-1557(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA BLOECKER H., BRANDT P.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 460-833 FROM N.A.
RA BENTON B.K., PLUMP S.D., ROOS J., LENNARZ W.J., CROSS F.R.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-100.
RX MEDLINE; 97446148.
RA TAYLOR I.A., TREIBER M.K., OLIVI L., SMERDON S.J.;
RT "The X-ray structure of the DNA-binding domain from the Saccharomyces
RT cerevisiae cell-cycle transcription factor Mbpl at 2.1-A
RT resolution";
RL J. MOL. BIOL. 272:1-8(1997).
CC -1- FUNCTION: BINDS TO MCB ELEMENTS (MLU I CELL CYCLE BOX) FOUND IN
CC THE PROMOTER OF MOST DNA SYNTHESIS GENES. TRANSCRIPTIONAL
CC ACTIVATION BY MBF HAS AN IMPORTANT ROLE IN THE TRANSITION FROM G1
CC TO S PHASE. IT MAY HAVE A DUAL ROLE IN THAT IT BEHAVES AS AN
CC ACTIVATOR OF TRANSCRIPTION AT THE G1-S BOUNDARY AND AS A REPRESSOR
CC DURING OTHER STAGES OF THE CELL CYCLE.
CC -1- SUBUNIT: MBF CONTAINS SWI6 AND MBP1.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS 2 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X74158; G402793; -
DR EMBL; X74158; G402793; -
DR EMBL; X74158; G402793; -
DR EMBL; U19608; G639493; -
DR PIR; A47528; A47528.
DR PIR; S37404; S37404.
DR PDB; 1MB1; 29-JUL-98.
DR SGD; L0001032; MBP1.
DR PFAM; PF00023; ank; 2.
KW TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN; DNA-BINDING;
KW ANK REPEAT; REPEAT; 3D-STRUCTURE.
FT DNA_BIND 1 94
FT DOMAIN 388 538 2 X ANK MOTIF REPEATS.
FT REPEAT 388 420 ANK MOTIF 1.
FT REPEAT 506 538 ANK MOTIF 2.
SQ SEQUENCE 833 AA; 93907 MW; 20FAB3BD CRC32;

Query Match 64.6%; Score 51; DB 1; Length 833;
Best Local Similarity 40.0%; Pred. No. 3.90e+00;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 433 MRSSLFHSKY 442
| | | | |
QY 365 LKTPLFHFK 374

```

RESULT 9
ID SEFL_YEAST STANDARD; PRT: 1057 AA.
AC P34228;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SUPPRESSOR PROTEIN SEFL
GN SEFL OR YBL066C OR YBL0501 OR YBL0526.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE OF 1-788 FROM N.A.
RC STRAIN-S288C;
RX MEDLINE; 94203266.
RA SCHERENS B., EL BAKKOURY M., VIERENDELS F., DUBOIS E., MESSENGUY F.;
RT "Sequencing and functional analysis of a 32,560 bp segment on the
RT left arm of yeast chromosome II. Identification of 26 open reading
RT frames, including the KIP1 and SECL7 genes.";
RL YEAST 9:1355-1371(1993).
RN [2]
RP SEQUENCE OF 285-1057 FROM N.A.
RC STRAIN-S288C;
RA CONTRERAS R., FIERES W., LOGGHE M., MOLEMANS F.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP GENE NAME.
RX MEDLINE; 98144791.
RA GROOM K.R., HEYMAN H.C., STEFFEN M.C., HAWKINS L., MARTIN N.C.;
RT "Kluyveromyces lactis SEFL and its Saccharomyces cerevisiae homologue
RT bypass the unknown essential function, but not the mitochondrial
RT kinase P function, of the S. cerevisiae RPM2 gene.";
RL YEAST 14:77-87(1998).
CC -!- FUNCTION: SEEMS TO BE INVOLVED IN THE SPOULATION PROCESS.
CC -!- SUPPRESSES THE LETHAL PHENOTYPE OF RPM2 DELETION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z33261; G313734; -
DR EMBL; Z35827; G536103; -
DR PIR; S45801; S45801.
DR SGD; L0003318; SEFL.
DR PROSITE; PS00463; ZN2_CYS6_FUNGAL_1; 1.
DR PROSITE; PS50048; ZN2_CYS6_FUNGAL_2; 1.
DR PFAM; PF00172; ZN2_CYS6; 1.
KW TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN; ZINC;
KW METAL-BINDING; SPOULATION.
FT DNA_BIND 57 87 ZN(2)-CYS(6), FUNGAL-TYPE.
SQ SEQUENCE 1057 AA; 118192 MW; 1B263749 CRC32;

Query Match 54.6%; Score 51; DB 1; Length 1057;
Best Local Similarity 66.7%; Pred. No. 3.90e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1038 RTPLFRSIY 1046
QY 366 KTLPLFSKY 374

RESULT 10
ID YA45_SCHPO STANDARD; PRT: 1583 AA.
AC Q09725;

DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 180.2 KD PROTEIN C31A2.05C IN CHROMOSOME I.
GN SPAC31A2.05C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z50113; G914883; -
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 319 339 POTENTIAL.
FT TRANSMEM 633 653 POTENTIAL.
FT TRANSMEM 764 784 POTENTIAL.
FT TRANSMEM 1014 1034 POTENTIAL.
FT TRANSMEM 1216 1236 POTENTIAL.
FT TRANSMEM 1452 1472 POTENTIAL.
SQ SEQUENCE 1583 AA; 180203 MW; 9617C75D CRC32;

Query Match 64.6%; Score 51; DB 1; Length 1583;
Best Local Similarity 54.5%; Pred. No. 3.90e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 446 LKAPLFQTLQY 456
QY 365 LKTLPLFHS-KY 374

RESULT 11
ID CBP3_YEAST STANDARD; PRT: 335 AA.
AC P21560;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CBP3 PROTEIN PRECURSOR.
GN CBP3 OR YPL215W.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89291856.
RA WU M., TZAGOLOFF A.;
RT "Identification and characterization of a new gene (CBP3) required
RT for the expression of yeast coenzyme QH2-cytochrome c reductase.";
RL J. BIOL. CHEM. 264:11122-11130(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA RIEGER M., MUELLER-AUER S., SCHAEFER M.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: REQUIRED FOR THE ASSEMBLY OF YEAST COENZYME
CC QH-2-CYTOCHROME C REDUCTASE.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; J04830; G171171; -;
DR EMBL; Z73571; E247067; -;
DR PIR; A34290; A34290;
DR SGD; L0000226; CBP3.
KW MITOCHONDRION; TRANSMEMBRANE; TRANSIT PEPTIDE.
FT CHAIN 1 38 MITOCHONDRION (POTENTIAL).
FT TRANSIT 39 335 CBP3 PROTEIN.
FT TRANSMEM 152 169 POTENTIAL.
FT MOD_RES 335 AA; 39083 MW; 85CFAAF8 CRC32;
SQ SEQUENCE 335 AA; 63.3%; Score 50; DB 1; Length 335;

Query Match 63.3%; Score 50; DB 1; Length 335;
Best Local Similarity 77.8%; Pred. No. 6.34e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 73 KIPLAHISKY 81
| | | | |
Qy 366 KTLPLFHSKY 374

RESULT 12
ID DSPA_SYNY3 STANDARD; PRT; 663 AA.
AC P20169; Q55232;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DRUG SENSORY PROTEIN A (EC 2.7.3.-).
GN DSPA OR DFR OR SLL0698.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96081505.
RA BARTSEVICH V.V., SHESTAKOV S.V.;
RT "The dspA gene product of the cyanobacterium Synechocystis sp. strain
PCC 6803 influences sensitivity to chemically different growth
inhibitors and has amino acid similarity to histidine protein
kinases".
RT MICROBIOLOGY 141:2915-2920(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97051201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K.,
RA OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,
RA YAMADA M., YASUDA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions".
RL DNA RES. 3:109-136(1996).
RN [3]
RP SEQUENCE OF 1-9 FROM N.A.
RX MEDLINE; 89034300.
RA REILLY P., HULMES J.D., PAN Y.C.E., NELSON N.;
RT "Molecular cloning and sequencing of the psad gene encoding subunit
II of photosystem I from the cyanobacterium, Synechocystis sp. PCC
6803".
RL J. BIOL. CHEM. 263:17658-17662(1988).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- MUTATIONS IN DSPA RESULTS IN RESISTANCE TO HERBICIDES DIFUNON,
CC DIURON AND CALMODULIN ANTAGONISTS CHLORPROMAZINE AND
CC TRIFLUOPERAZINE.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X72856; G288420; -;
DR EMBL; D90899; D1017420; -;
DR PIR; J04195; -; NOT_ANNOTATED_CDS.
DR PIR; D32124; D32124.
DR PFAM; PF00512; signal; 1.
DR PFAM; PF00672; DUF5; 1.
DR PFAM; PF00989; PAS; 1.
KW HERBICIDE RESISTANCE; SENSORY TRANSDUCTION; TRANSFERASE; KINASE;
KW TRANSMEMBRANE; PHOSPHORYLATION.
FT TRANSMEM 32 52 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT MOD_RES 432 432 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 663 AA; 74536 MW; 43858F25 CRC32;

Query Match 63.3%; Score 50; DB 1; Length 663;
Best Local Similarity 66.7%; Pred. No. 6.34e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 434 LRTPLFNK 442
|:|:|:|:|
Qy 365 LKTPLFHSK 373

RESULT 13
ID COXM_HUMAN STANDARD; PRT; 80 AA.
AC P24311;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE CYTOCHROME C OXIDASE POLYPEPTIDE VIIB PRECURSOR (EC 1.9.3.1).
GN COX7B.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
RN PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93176819.
RA SADLOCK J.E., LIGHTOWLERS R.N., CAPALDI R.A., SCHON E.A.;
RT "Isolation of a cDNA specifying subunit VIIB of human cytochrome c
oxidase".
RL BIOCHIM. BIOPHYS. ACTA 1172:223-225(1993).
RN [2]
RP SEQUENCE OF 25-43.
RX TISSUE=HEART;
RX MEDLINE; 92111565.
RA VAN KUILENBURG A.B.P., VAN BEEUMEN J.J., VAN DER MEER N.M.,
RA MUIJSERS A.O.;
RT "Subunits VIIa,b,c of human cytochrome c oxidase. Identification of
both 'heart-type' and 'liver-type' isoforms of subunit VIIa in human
heart".
RL EUR. J. BIOCHEM. 203:193-199(1992).
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
CC CHAINS OF CYTOCHROME C OXIDASE. THE TERMINAL OXIDASE IN
CC MITOCHONDRIAL ELECTRON TRANSPORT.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z14244; G30151; -;
DR PIR; S20289; S20289.
DR PIR; S29856; S29856.

KW OXIDOREDUCTASE; MITOCHONDRION; TRANSIT PEPTIDE.
FT TRANSIT 1 24 MITOCHONDRION.
FT CHAIN 25 80 CYTOCHROME C OXIDASE POLYPEPTIDE VIIB.
FT CONFLICT 29 29 R>P (IN REF. 2).
SQ SEQUENCE 80 AA: 9160 MW; 288DE71F CRC32;
Query Match 62.0%; Score 49; DB 1; Length 80;
Best Local Similarity 66.7%; Pred. No. 1.02e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 29 RTPDFHDKY 37
: || || ||
QY 366 KTLPLFHSKY 374
RESULT 14
ID YBF5_YEAST STANDARD; PRT: 149 AA.
AC P38730;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 17.6 KD PROTEIN IN CBP2 5'REGION.
GN YH041W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE; 94378003.
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,
RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
RA NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,
RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
RA VAUDIN M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII.";
RL SCIENCE 265:2077-2082(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U11593; G2289870; -
DR PIR; S48927; S48927.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 149 AA: 17554 MW; E04E7779 CRC32;

Query Match 62.0%; Score 49; DB 1; Length 149;
Best Local Similarity 60.0%; Pred. No. 1.02e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 56 IKTPFIASY 65
: || || || : |
QY 365 LKTPLEFSKY 374

RESULT 15
ID YBF5_YEAST STANDARD; PRT: 418 AA.
AC P34220;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 47.4 KD PROTEIN IN PTC3-SAS3 INTERGENIC REGION.
GN YBL055C OR YBL0512 OR YBL0511.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;

OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE; 94205266.
RA SCHERENS B., EL BAKKOURY M., VIERENDEELS F., DUBOIS E., MESSENGUY F.;
RT "Sequencing and functional analysis of a 32,560 bp segment on the
RT left arm of yeast chromosome II. Identification of 26 open reading
RT frames, including the KIP1 and SEC17 genes.";
RL YEAST 9:1355-1371(1993).
CC -!- SIMILARITY: BELONGS TO THE UPF0006 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 223261; G313743; -
DR EMBL; Z35816; G536084; -
DR PIR; S39833; S39833.
DR PIR; S37334; S37334.
DR PROSITE; PS01137; UPF0006_1; FALSE_NEG.
DR PROSITE; PS01090; UPF0006_2; 1.
DR PROSITE; PS01091; UPF0006_3; 1.
DR PFAM; PF01026; UPF0006; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 418 AA: 47390 MW; A09BF15B CRC32;
Query Match 62.0%; Score 49; DB 1; Length 418;
Best Local Similarity 50.0%; Pred. No. 1.02e+01;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Db 45 LTDPWFHGIY 54
: || || || : |
QY 365 LKTPLEFSKY 374

Search completed: Thu Jul 8 18:50:28 1999
Job time : 10 secs.

THIS PAGE BLANK (USPTO)

W P E R L A (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:51:21 1999; Maspar time 5.89 Seconds
Tabular output not generated. 92.739 Million cell updates/sec

Title: >US-09-041-236-2
Description: (365-374) from US09041236.pap (22 of 45)
Perfect Score: 79
Sequence: 1 LKTPLEFSKY 10

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptribl9
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 24.252; Variance 29.611; scale 0.819

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	79	100.0	666	4	SEMAPHORIN L.	3.55e-06
2	56	70.9	209	10	F6N15.9 PROTEIN.	9.68e-01
3	56	70.9	325	5	F5GA6.1 PROTEIN (FRAGM	9.68e-01
4	56	70.9	697	5	CODED FOR BY C. ELEGAN	9.68e-01
5	53	67.1	210	10	HYPOTHETICAL 23.5 KD P	4.15e+00
6	53	67.1	418	10	PUTATIVE TETRACYCLINE	4.15e+00
7	53	67.1	572	3	CEPHALOSPORIN ESTERASE	4.15e+00
8	52	65.8	361	2	MANNOSYL TRANSFERASE.	6.65e+00
9	52	65.8	364	11	NKX6.1.	6.65e+00
10	52	65.8	546	5	F53G2.7 PROTEIN.	6.65e+00
11	51	64.6	255	10	PUTATIVE ASPARTIC PROT	1.06e+01
12	51	64.6	422	4	CYTOKINE-LIKE FACTOR-1	1.06e+01
13	51	64.6	506	10	ASPARTIC PROTEASE.	1.06e+01
14	51	64.6	509	10	ASPARTIC PROTEINASE.	1.06e+01
15	51	64.6	509	10	CYPROSIN.	1.06e+01
16	51	64.6	585	5	F42A6.3 PROTEIN.	1.06e+01
17	51	64.6	652	3	HYPOTHETICAL 73.8 KD P	1.06e+01
18	51	64.6	701	5	HEXAMERIN LSP-2.	1.06e+01
19	50	63.3	75	11	DYNEIN, AXON, HEAVY CH	1.68e+01
20	50	63.3	254	4	NON-FUNCTIONAL FOLATE	1.68e+01

21 50 63.3 349 2 025802 ADP-HEPTOSE-LPS HEPTOS 1.68e+01
22 50 63.3 352 8 063697 RIBOSOMAL PROTEIN S4. 1.68e+01
23 50 63.3 669 10 081752 HYPOTHETICAL 75.2 KD P 2.64e+01
24 49 62.0 191 4 000588 RADICAL FRINGE (FRAGME 2.64e+01
25 49 62.0 221 5 001256 T20D3.5 PROTEIN. 2.64e+01
26 49 62.0 223 2 067786 ENOLASE-PHOSPHATASE E- 2.64e+01
27 49 62.0 277 2 086305 FTSZ PROTEIN (FRAGMENT 2.64e+01
28 49 62.0 288 2 P71025 CZCD (CZCD). 2.64e+01
29 49 62.0 293 2 P74944 COMPONENT OF SODIUM-DR 2.64e+01
30 49 62.0 332 11 009009 RADICAL FRINGE PROTEIN 2.64e+01
31 49 62.0 365 11 035762 HOMEODOMAIN PROTEIN NK 2.64e+01
32 49 62.0 367 4 P78426 NK HOMEBOX PROTEIN. 2.64e+01
33 49 62.0 372 13 012970 RADICAL FRINGE. 2.64e+01
34 49 62.0 372 13 012972 RADICAL FRINGE. 2.64e+01
35 49 62.0 393 5 Q18315 CODED FOR BY C. ELEGAN 2.64e+01
36 49 62.0 425 14 064796 EARLY 49 KDA PROTEIN. 2.64e+01
37 49 62.0 572 5 076835 T12A2.15 PROTEIN. 2.64e+01
38 49 62.0 662 5 017110 F39E9.4 PROTEIN. 2.64e+01
39 49 62.0 744 4 015404 CAGF28 (FRAGMENT). 2.64e+01
40 49 62.0 837 5 Q22413 HYPOTHETICAL 95.4 KD P 2.64e+01
41 49 62.0 881 5 016386 K12B6.1 PROTEIN. 2.64e+01
42 49 62.0 932 5 Q93562 F25D7.4 PROTEIN. 2.64e+01
43 49 62.0 954 5 045527 F47H4.2 PROTEIN. 2.64e+01
44 49 62.0 1568 3 074415 HYPOTHETICAL 179.2 KD 2.64e+01
45 49 62.0 3871 5 Q20911 ZC116.3 PROTEIN. 2.64e+01

ALIGNMENTS

RESULT 1
ID 075326 PRELIMINARY; PRT; 666 AA.
AC 075326;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN L.
GN SEMAL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses".
RL GENOMICS 51:340-350(1998).
DR EMBL: AF030698; G3523115; --
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 79; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 3.55e-06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 397 LKTPLEFSKY 406
QY 365 LKTPLEFSKY 374
|||||||

RESULT 2
ID 081309 PRELIMINARY; PRT; 209 AA.
AC 081309;

DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F6N15.9 PROTEIN.
GN F6N15.9.

OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT:
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AL021711; E1249630; -
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 210 AA; 23512 MW; 11A9D729 CRC32;

Query Match 67.1%; Score 53; DB 10; Length 210;
Best Local Similarity 60.0%; Pred. No. 4.15e+00;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 63 LKTLFHDFT 72
||| ||| :
Qy 365 LKTLFHSKY 374

RESULT 6
ID O22995 PRELIMINARY; PRT; 418 AA.
AC O22995;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PUTATIVE TETRACYCLINE TRANSPORTER-LIKE PROTEIN.
GN F6P23.2.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA HASEGAWA A.E., SCHUTZ K., DE LA BASTIDE M., LODHI M., GNOJ L.,
RA GOTESMAN T., GRANAT S., HAMEED A., KAPLAN N., SHOHDY N.,
RA VAN KEUREN K., MARRA M., JOHNSON A.F., PARNELL L., DEDHIA N.,
RA MARTENSEN R., MCCOMBIE W.R.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AC002354; G2264382; -
SQ SEQUENCE 418 AA; 45828 MW; 47E89EE8 CRC32;

Query Match 67.1%; Score 53; DB 10; Length 418;
Best Local Similarity 60.0%; Pred. No. 4.15e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 216 LKTVFHKY 225
:|: ||| :
Qy 365 LKTLFHSKY 374

RESULT 7
ID O42728 PRELIMINARY; PRT; 572 AA.
AC O42728;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CEPHALOSPORIN ESTERASE PRECURSOR.
OS RHODOSPORIDIUM TORULOIDES (YEAST).
OC EUKARYOTA; FUNGI; BASIDIOMYCOTA; UREDINIOMYCETES; SPORIDIALES;
OC SPORIDIACEAE; RHODOSPORIDIUM.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98059473.
RA POLITINO M., TONZI S.M., BURNETT W.V., ROMANCIK G., USHER J.J.;
RT "Purification and characterization of a cephalosporin esterase from
Rhodosporidium toruloides."
RL APPL. ENVIRON. MICROBIOL. 63:4807-4811(1997).
DR EMBL; AF025410; G2731568; -
SQ SEQUENCE 572 AA; 61315 MW; 5FD70284 CRC32;

Query Match 67.1%; Score 53; DB 3; Length 572;
Best Local Similarity 75.0%; Pred. No. 4.15e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 256 LKKPLFHA 263

Qy 365 LKTLFHS 372
|| |||||

RESULT 8
ID Q54128 PRELIMINARY; PRT; 361 AA.
AC Q54128; Q53988;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MANNOSYL TRANSFERASE.
GN RFBO.
OS SALMONELLA TYPHIMURIUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=GROUP E1, SEROVAR ANATUM, M32;
RX MEDLINE; 92201631.
RA WANG L., ROMANA L.K., REEVES P.R.;
RT "Molecular analysis of a Salmonella enterica group E1 rfb gene
cluster: O antigen and the genetic basis of the major polymorphism.";
RL GENETICS 130:429-443(1992).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=GROUP E1, SEROVAR ANATUM, M32;
RX MEDLINE; 93273708.
RA LIU D., HAASE A.M., LINDQVIST L., LINDBERG A.A., REEVES P.R.;
RT "Glycosyl transferases of O-antigen biosynthesis in Salmonella
enterica: identification and characterization of transferase genes of
groups B, C2, and E1".
RL J. BACTERIOL. 175:3408-3413(1993).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=GROUP E1, SEROVAR ANATUM, M32;
RX MEDLINE; 93166587.
RA REEVES P.R.;
RT "Evolution of Salmonella O antigen variation by interspecific gene
transfer on a large scale."
RL TRENDS GENET. 9:17-22(1993).
[4]
RP SEQUENCE OF 1-102 FROM N.A.
RC STRAIN=GROUP D2, SEROVAR STRASBOURG;
RX MEDLINE; 94292466.
RA XIANG S., HOBBS M., REEVES P.R.;
RT "Molecular analysis of the rfb gene cluster of a group D2 Salmonella
enterica strain: evidence for its origin from an insertion
sequence-mediated recombination event between group E and D1
strains".
RL J. BACTERIOL. 176:4357-4365(1994).
DR EMBL; X60665; G46989; -
DR EMBL; U04165; G529144; -
KW TRANSFERASE.
SQ SEQUENCE 361 AA; 40673 MW; 8A08656B CRC32;

Query Match 65.8%; Score 52; DB 2; Length 361;
Best Local Similarity 60.0%; Pred. No. 6.65e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 78 LKGILFNKY 87
|| |||||
Qy 365 LKTLFHSKY 374

RESULT 9
ID Q60554 PRELIMINARY; PRT; 364 AA.
AC Q60554;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE NKX6.1.
GN NKX6.1.
OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; CRICETINAE; MESOCRICETUS.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 95083670.
RX TISSUE-PANCREATIC ISLET;
RA RUDNICK A., LING T.Y., ODAGIRI H., RUTTER W.J., GERMAN M.S.;
RT "Pancreatic beta cells express a diverse set of homeobox genes.";
RL PROC. NATL. ACAD. SCI. U.S.A. 91:12203-12207(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR ENBL; X81409; G587467; -.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PFAM; PF00046; homeobox; 1.
KW HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN.
SQ SEQUENCE 364 AA; 37646 MW; 92CD6E89 CRC32;

Query Match 65.8%; Score 52; DB 11; Length 364;
Best Local Similarity 50.0%; Pred. No. 6.65e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 31 MKTPLYPATY 40
QY 365 LKTPLFHSKY 374
:||||: :|

RESULT 10
ID O17245 PRELIMINARY; PRT; 546 AA.
AC O17245;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F53G2.7 PROTEIN.
GN F53G2.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BECKER M., WOHLDMANN P., BIEWALD T.;
RL SUBMITTED (SEP-1997) TO ENBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (SEP-1997) TO ENBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR ENBL; AF025464; G2429484; -.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
KW ZINC-FINGER.
SQ SEQUENCE 546 AA; 65219 MW; 2A9AFEE5 CRC32;

Query Match 65.8%; Score 52; DB 5; Length 546;
Best Local Similarity 66.7%; Pred. No. 6.65e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 470 KMOLFSKF 478
QY 366 KTFPLFSKY 374
:||||: :|

RESULT 11
ID Q43406 PRELIMINARY; PRT; 255 AA.
AC Q43406;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE ASPARTIC PROTEASE (FRAGMENT).
OS BRASSICA OLERACEA (CAULIFLOWER).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; BRASSICA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ALPHA PAL; TISSUE=ROOT TIPS;
RA FUJIKURA Y., KARSEN C.M.;
RL SEED SCI. RES. 5:177-181(1995).
DR ENBL; X80067; G510880; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PFAM; PF00026; asp; 1.
KW PROTEASE; HYDROLASE; ASPARTYL PROTEASE.
FT NON_TER 255
SQ SEQUENCE 255 AA; 27912 MW; D35F5289 CRC32;

Query Match 64.6%; Score 51; DB 10; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.06e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 120 LFHSKY 125
QY 369 LFHSKY 374
:||||: :|

RESULT 12
ID O75462 PRELIMINARY; PRT; 422 AA.
AC O75462;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CYTOKINE-LIKE FACTOR-1 PRECURSOR.
GN CLF-1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA ELSON G.C.A., GRABER P., LOSBERGER P., HERREN S., GRETENER D.,
RA MENOUD L.N., WELLS T.N.C., KOSCO-VILBOIS M.H., GAUCHAT J.F.;
RT "CLF-1, a Novel Soluble Protein Shares Homology with Members of the
RT Cytokine Type-I Receptor Family";
RL J. IMMUNOL. 0:0-0(1998).
DR ENBL; AF059293; G3372627; -.
KW SIGNAL.
FT SIGNAL. 1 37 POTENTIAL.
FT CHAIN 38 422 CYTOKINE-LIKE FACTOR-1.
SQ SEQUENCE 422 AA; 46301 MW; 877F9BC9 CRC32;

Query Match 64.6%; Score 51; DB 4; Length 422;
Best Local Similarity 60.0%; Pred. No. 1.06e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 263 LKDFLFQAKY 272
QY 365 LKTPLFHSKY 374
:||||: :|

RESULT 13
ID Q39311 PRELIMINARY; PRT; 506 AA.
AC Q39311;

DT 01-NOV-1996 (TREMREL. 01, CREATED)
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
 DE ASPARTIC PROTEINASE.
 OS BRASSICA NAPUS (RAPE).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; BRASSICA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94012630.
 RA D'HONT K., BOSCH D., VAN DAMME J., GOETHALS M., VANDEKERCKHOVE J.,
 RA KREBBERS E.;
 RT "An aspartic proteinase present in seeds cleaves Arabidopsis 2 S
 albumin precursors in vitro";
 RL J. BIOL. CHEM. 268:20884-20891(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA D'HONT K., STACK S., GUTTERIDGE S., VANDEKERCKHOVE J., KREBBERS E.,
 RA GAL S.;
 RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U55032; G1326165; -;
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 DR PFAM; PF00026; asp; 1.
 DR MENDEL; 7339; BRANA; 1036;1.
 KW PROTEASE; HYDROLASE; ASPARTYL PROTEASE.
 SQ SEQUENCE 506 AA; 54295 MW; 85C216D1 CRC32;

Query Match 64.6%; Score 51; DB 10; Length 506;
 Best Local Similarity 100.0%; Pred. No. 1.06e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 120 LFHSKY 125
 |
 Qy 369 LFHSKY 374

RESULT 14
 ID Q96383 PRELIMINARY; PRT; 509 AA.
 AC Q96383;
 DT 01-FEB-1997 (TREMREL. 02, CREATED)
 DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
 DE ASPARTIC PROTEINASE.
 OS CENTAUREA CALCITRAPA.
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
 OC ASTERIDAE; ASTERALES; ASTERACEAE; CENTAUREA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FLOWERS;
 RA DOMINGOS A., XUE Z.T., PAIS M.S., CLEMENTE A., BRODELIUS P.E.;
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; Y09123; E280545; -;
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 DR PFAM; PF00026; asp; 1.
 DR MENDEL; 11619; CENCA; 1036;1.
 KW HYDROLASE; ASPARTYL PROTEASE.
 FT CHAIN 69 509 ASPARTIC PROTEINASE.
 SQ SEQUENCE 509 AA; 55264 MW; 4718B3E9 CRC32;

Query Match 64.6%; Score 51; DB 10; Length 509;
 Best Local Similarity 100.0%; Pred. No. 1.06e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 123 LFHSKY 128
 |
 Qy 369 LFHSKY 374

RESULT 15
 ID Q39476 PRELIMINARY; PRT; 509 AA.
 AC Q39476;

DT 01-NOV-1996 (TREMREL. 01, CREATED)
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
 DE CYPROSIN.
 OS CYNARA CARDUNCULUS (CARDOON).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
 OC ASTERIDAE; ASTERALES; ASTERACEAE; CYNARA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FLOWER;
 RA PIETRZAK M., BRODELIUS P., PAIS M.S.;
 RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; X81984; G556819; -;
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 DR PFAM; PF00026; asp; 1.
 DR MENDEL; 11708; CYNCA; 1036;2.
 KW HYDROLASE; ASPARTYL PROTEASE.
 SQ SEQUENCE 509 AA; 55176 MW; 50ABCA04 CRC32;

Query Match 64.6%; Score 51; DB 10; Length 509;
 Best Local Similarity 100.0%; Pred. No. 1.06e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 123 LFHSKY 128
 |
 Qy 369 LFHSKY 374

Search completed: Thu Jul 8 18:51:36 1999
 Job time : 15 secs.

THIS PAGE BLANK (USPTO)

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Thu Jul 8 18:54:23 1999; MasPar time 11.21 Seconds
Tabular output not generated.
17.075 Million cell updates/sec

Title: >US-09-041-236-2
Description: (412-420) from US09041236.pep (23 of 45)
Perfect Score: 63
Sequence: 1 EOEHSFAFN 9
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 15.390; Variance 50.525; scale 0.305

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	46	73.0	419	17	R97618 Human carboxypeptidase	2.06e+02
2	46	73.0	419	20	W01508 Human pancreatic carb	2.06e+02
3	46	73.0	419	20	W01504 Human pancreatic carb	2.06e+02
4	46	73.0	419	20	W01504 Wild-type human pancr	2.06e+02
5	46	73.0	419	20	W01509 Human pancreatic carb	2.06e+02
6	44	69.8	9	29	W42463 Mouse anti-human IL5	3.26e+02
7	44	69.8	9	18	R98487 Mouse anti-human IL5	3.26e+02
8	44	69.8	105	18	R98497 MAb 2E3 light chain v	3.26e+02
9	44	69.8	105	18	R98495 MAb 2F2 light chain v	3.26e+02
10	44	69.8	105	29	W42454 Mouse anti-human IL5	3.26e+02
11	44	69.8	105	29	W42456 Mouse anti-human IL5	3.26e+02
12	44	69.8	232	6	R29716 Guinea pig MBP-1	3.26e+02
13	44	69.8	233	29	W40075 Guinea pig eosinophil	3.26e+02
14	43	68.3	17	1	P82495 Cyclic peptide with A	4.10e+02
15	42	66.7	16	33	W62380 Antithrombotic peptid	5.13e+02
16	42	66.7	591	29	W41418 Paxillin beta isoform	5.13e+02

17	42	66.7	605	29	W41419 Paxillin gamma isofo	5.13e+02
18	42	66.7	1085	18	R95607 RRP3 telomerase-assoc	5.13e+02
19	41	65.1	60	33	W62371 Antithrombotic peptid	6.42e+02
20	41	65.1	152	29	W48272 Human ninjurin 1	6.42e+02
21	41	65.1	223	33	W61218 Streptococcus pneumon	6.42e+02
22	41	65.1	308	22	W20274 H. pylori secreted or	6.42e+02
23	41	65.1	316	22	W20812 H. pylori secreted or	6.42e+02
24	41	65.1	457	35	W59867 Amino acid sequence o	6.42e+02
25	41	65.1	489	19	W01123 Phytoene dehydrogenas	6.42e+02
26	41	65.1	489	3	R13984 Phytoene dehydrogenas	6.42e+02
27	41	65.1	489	18	W00174 Phytoene dehydrogenas	6.42e+02
28	41	65.1	489	25	W32475 Erwinia herbicola phy	6.42e+02
29	41	65.1	489	25	W32472 Erwinia herbicola phy	6.42e+02
30	41	65.1	489	19	W01122 Phytoene dehydrogenas	6.42e+02
31	41	65.1	489	3	R13985 Phytoene dehydrogenas	6.42e+02
32	41	65.1	489	18	W00173 Phytoene dehydrogenas	6.42e+02
33	41	65.1	732	25	W26642 Human RECK cancer-inh	6.42e+02
34	41	65.1	777	6	R30077 Pisum sativum L PAL	6.42e+02
35	41	65.1	1003	28	W35003 Thermotoga maritima e	6.42e+02
36	41	65.1	2324	16	R92778 Human fibronectin	6.42e+02
37	41	65.1	2327	3	R15468 Human fibronectin	6.42e+02
38	41	65.1	2327	3	P70373 Human fibronectin gen	6.42e+02
39	41	65.1	2386	34	W63171 Amino acid sequence o	6.42e+02
40	41	65.1	2446	11	R60021 Fibrinogen-alpha	6.42e+02
41	41	65.1	3135	11	R57474 P. faiciparum transmi	6.42e+02
42	40	63.5	452	39	W87889 Protein encoded by th	8.01e+02
43	40	63.5	648	22	W17045 Mouse c-raf 1 protein	8.01e+02
44	40	63.5	993	7	R37503 Human flk-2	8.01e+02
45	40	63.5	2639	39	W73476 Grapevine leafroll vi	8.01e+02

ALIGNMENTS

RESULT 1
ID R97618 standard; Protein; 419 AA.
AC R97618:
DE Human carboxypeptidase Al.
KW Gene therapy; gene directed enzyme produg therapy; GDEPT;
KW virus directed enzyme produg therapy; VDEPT; produg activation;
KW cytotoxic; cytostatic; cancer; tumour; retrovirus; vector;
KW carboxypeptidase Al; HCPAL.
OS Homo sapiens.
PN W09616179-Al.
PD 30-MAY-1996.
PR 20-NOV-1995; G02716.
PR 18-NOV-1994; GB-023367.
PA (WELL) WELLCOME FOUND LTD.
PI Dev IK, Moore JT, Ohmstede C;
DR WPI; 96-268615/27.
DR N-PSDB; T29217.
PT Molecular chimera for use in enzyme gene therapy - is activated in
PT a target cell to express a secretable enzyme which cleaves a produg
PT outside the cell into a cytotoxic or cytostatic agent
PS Example 2; Page 43-45; 73pp; English.
CC The amino acid sequence (R97618) of human pancreatic
CC carboxypeptidase Al (HCPAL) was deduced from a cDNA clone (T29217)
CC isolated from a human pancreas cDNA library. Mutagenesis of the cDNA
CC provides clones coding for mutant enzymes, pref. 268G-HCPAL. A
CC molecular chimera can be produced that comprises a transcriptional
CC regulatory sequence activated in a targeted mammalian cell, e.g.
CC tumour cells, linked to a sequence coding for a signal peptide and
CC 268G-HCPAL. Upon expression, 268G-HCPAL catalyzes extracellular
CC conversion of a produg into a cytotoxic or cytostatic agent. Gene
CC or virus directed enzyme produg therapy is obtd.
SQ Sequence 419 AA;

Query Match 73.0%; Score 46; DB 17; Length 419;
Best Local Similarity 75.0%; Pred. No. 2.06e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 100 eeqgmfaf 107
|||: |||


```
Db 100 eeqmfaf 107
    |||: |||
QY 412 EOEHSFAP 419

RESULT 5
ID W01509 standard; Protein; 419 AA.
AC W01509;
DE Human pancreatic carboxypeptidase 1 variant (T268G,A).
KW HCPAL; pancreatic; anti-tumour therapy; anti-neoplastic; prodrug;
KW ADEPT; antibody directed enzyme prodrug therapy; carboxypeptidase A.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_difference 378
FT /label= Ala, Gly
FT /note= "wild-type Thr substituted with Gly; claim 34 "

PN W09513095-A2.
PD 18-MAY-1995.
PF 11-NOV-1994; G02483.
PR 12-NOV-1993; GB-023429.
PA (WELL ) WELLCOME FOUND LTD.
PI Blumenkopf TA, Cory M, Smith GK.
DR WPI; 95-193907/25.
PT New prodrugs, and enzyme-targeting molecule conjugates - useful in
PT antibody-directed enzyme prodrug therapy of e.g. viral infections
PS Claim 34; Page -; 247pp; English.
CC A new conjugate for anti-tumour therapy comprises a targeting moiety
CC linked to a mutated enzyme which can transform an anti-neoplastic
CC prodrug. The system is based on antibody directed enzyme prodrug therapy
CC (ADEPT) using a non-naturally occurring mutant form of a host enzyme.
CC The enzyme precursor is refractory to endogenous catalysis by the wild-
CC type form of the enzyme. The mutated enzyme has markedly reduced
CC immunogenicity compared to non-host enzymes. Also because the mutated
CC enzyme is non-naturally occurring, non-selective triggering of prodrug
CC activation by natural flora or host enzymes is reduced. The enzyme is
CC pref. a hydrolase, a transferase, an oxido-reductase, an isomerase, a
CC lyase or a ligase, esp. human pancreatic carboxypeptidase A1 or 2 (HCPAL
CC or 2) possessing amino acid substitutions at one or more of residues 203,
CC 210, 242, 244, 250, 253, 255, 267, 268, 269 and/or 305, all other
CC residues being wild-type. The present sequence is a preferred variant of
CC HCPAL.
SQ Sequence 419 AA;


```

Query Match 73.0%; Score 46; DB 20; Length 419;
Best Local Similarity 75.0%; Pred. No. 2.06e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Db 100 eeqmfaf 107
    |||: |||
QY 412 EOEHSFAP 419

RESULT 6
ID W42463 standard; Peptide; 9 AA.
AC W42463;
DE Mouse anti-human IL5 Mab 2F2 and 2E3 light chain CDR2.
KW Neutralising antibody; monoclonal antibody; Mab; 2F2; 2E3; mouse;
KW interleukin-5; IL-5; human; eosinophil; asthma; allergic rhinitis;
KW atopic dermatitis; therapy; diagnosis; chimeric antibody;
KW humanised antibody; complementarity determining region; CDR.
OS Mus musculus.
PN W09748418-A1.
PD 20-DEC-1997.
PF 20-JUN-1997; U10769.
PR 21-JUN-1996; US-667769.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PI Appelbaum ER, Cook RM;
DR WPI; 98-062853/06.
PT Use of neutralising antibody to human interleukin-5 - that does not
PT block binding to receptor alpha chain, to treat e.g. asthma,
PT allergic rhinitis or atopic dermatitis
```

PS Example 4; Fig 7; 116pp; English.
CC This peptide comprises complementarity determining region 2 (CDR2)
CC of the light chain variable region of anti-human interleukin-5
CC (hIL-5) murine monoclonal antibodies (Mab) 2F2 and 2E3 (see W42454
CC and W42456). The invention provides antibodies, especially altered,
CC chimeric and humanised (see also W42465-66 and W42471-72),
CC which are characterised by hIL-5 specificity, neutralising activity
CC and affinity for hIL-5. The antibodies are prepared by conventional
CC hybridoma techniques, phage display combinatorial libraries,
CC immunoglobulin chain shuffling and humanisation techniques, e.g.
CC utilising the CDRs of mouse and rat anti-hIL-5 MAb. The products
CC are useful for treating hIL-5-mediated disorders such as asthma,
CC allergic rhinitis and atopic dermatitis, and can also be used in
CC the diagnosis of such conditions by measurement (e.g. by ELISA) of
CC endogenous hIL-5 levels.
SQ Sequence 9 AA;

Query Match 69.8%; Score 44; DB 29; Length 9;
Best Local Similarity 50.0%; Pred. No. 3.26e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1 qndhsfpf 8
 : |||: |
QY 412 EOEHSFAP 419

RESULT 7
ID R98487 standard; Peptide; 9 AA.
AC R98487;
DE 12-OCT-1996 (first entry)
DE Anti-IL-5 Mab 2F2/2E3 light chain variable region CDR3.
KW Antibody engineering; humanised antibody; chimeric antibody; Fab;
KW interleukin-5; IL-5; eosinophil; asthma; therapy; diagnosis;
KW complementarity determining region; CDR; light chain; VL;
KW monoclonal antibody; Mab.
OS Mus sp.
PN W09621000-A2.
PD 11-JUL-1996.
PF 22-DEC-1995; U17082.
PR 23-DEC-1994; US-363131.
PR 06-JUN-1995; US-470110.
PR 06-JUN-1995; US-467420.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Ames RS, Appelbaum ER, Chaiken IM, Cook RM, Gross MS;
PI Holmes SD, McMillan LJ, Theisen TW;
DR WPI; 96-333976/33.
PT New monoclonal antibody to human interleukin-5 - used to produce
PT products for the treatment and diagnosis of conditions associated
PT with excess eosinophil prodn., e.g. asthma etc.
PS Claim 12; Page 50; 120pp; English.
CC The complementarity determining regions (CDRs) for the VH region
CC of monoclonal antibody (Mab) 2B6 (see also R98478) are given in
CC R98480-82. Mab 2F2 VH (see also R98496), CDR1 and CDR2 are identical to
CC For Mab 2E3 VH (see also R98496), CDR1 and CDR2 are identical to
CC those of 2B6 and 2F2, but CDR3 has a different amino acid sequence
CC (R98483). The CDRs for the 2B6 VL region (see also R98479) are
CC given in R98484-86. For 2F2 VL (see also R98495) and 2E3 VL
CC (see also R98497), CDR1 and CDR2 are identical to CDR1 and CDR2
CC of 2B6, but CDR3 is different (R98487). 2B6, 2F2 and 2E3 are
CC murine anti-human interleukin-5 MABs. The CDRs can be used
CC in the construction of humanised antibodies (see also R98488-89)
CC and R98492-93) useful in the treatment of IL-5-mediated
CC conditions, e.g. asthma.
SQ Sequence 9 AA;

Query Match 69.8%; Score 44; DB 18; Length 9;
Best Local Similarity 50.0%; Pred. No. 3.26e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1 qndhsfpf 8
 : |||: |
QY 412 EOEHSFAP 419

Page 4

PT Use of neutralising antibody to human interleukin-5 - that does not
PT block binding to receptor alpha chain, to treat e.g. asthma,
PS allergic rhinitis or atopic dermatitis
PS Example 4; Fig 4; 116pp; English.
CC This polypeptide comprises the light chain variable region of
CC anti-human interleukin-5 (hIL-5) murine monoclonal antibody (MAB)
CC 2F2, which is produced by hybridoma cell line SK119-2F2.37.80.12
CC (ATCC HB 11781). In a claimed method for treating conditions
CC associated with excess eosinophil production, the improvement
CC comprises administering a neutralising MAB for hIL-5 which does not
CC block binding of hIL-5 to the alpha-chain of the hIL-5 receptor.
CC Also claimed are methods of detecting the absence/presence of
CC hIL-5 soluble receptor alpha-chain (Ralpha)/hIL-5 complexes,
CC and of screening for compounds that antagonise the binding of
CC Ralpha/hIL-5 or Ralpha/hIL-5/MAB complex with the hIL-5 receptor
CC beta-chain. Also disclosed are novel neutralising antibodies,
CC especially altered, chimeric or humanised antibodies (see
CC W42465-66, W42471-72), as well as vectors and transformed cells
CC for their expression. These products are useful for treating
CC hIL-5-mediated disorders such as asthma, allergic rhinitis and
CC atopic dermatitis, and can also be used in the diagnosis of such
CC conditions by measurement (e.g. by ELISA) of endogenous hIL-5
CC levels.
SQ Sequence 105 AA;

Query Match 59.8%; Score 44; DB 29; Length 105;
Best Local Similarity 50.0%; Pred. No. 3.26e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 87 qndhsfpf 94
: |||||
QY 412 EOEHSFAF 419

RESULT 11
ID W42456 standard; Protein; 105 AA.
AC W42456:
DT 22-JUN-1998 (first entry)
DE Mouse anti-human IL5 MAB 2E3 light chain variable region.
KW Neutralising antibody; monoclonal antibody; MAB; 2F2; mouse;
KW interleukin-5; IL-5; human; eosinophil; asthma; allergic rhinitis;
KW atopic dermatitis; therapy; diagnosis; chimeric antibody;
KW humanised antibody.
OS Mus musculus.
FH Key Location/Qualifiers
FT Region 16..32
FT /label= CDR1
FT /note= "complementarity determining region 1"
FT Region 48..54
FT /label= CDR2
FT /note= "complementarity determining region 2"
FT Region 87..95
FT /label= CDR3
FT /note= "complementarity determining region 3"
PW W09748418-A1.
PD 24-DEC-1997.
PF 20-JUN-1997; U10769.
PR 21-JUN-1996; US-667769.
PA (SMIX) SMITHKLINE BEECHAM CORP.
PI Appelbaum ER, Cook RM;
DR WPI; 98-062853/06.
DR N-PSDB; V03494.

PT Use of neutralising antibody to human interleukin-5 - that does not
PT block binding to receptor alpha chain, to treat e.g. asthma,
PT allergic rhinitis or atopic dermatitis
PS Example 4; Fig 6; 116pp; English.
CC This polypeptide comprises the light chain variable region of
CC anti-human interleukin-5 (hIL-5) murine monoclonal antibody (MAB)
CC 2E3, which is produced by hybridoma cell line SK119-2E3.39.40.2
CC (ATCC HB 11782). In a claimed method for treating conditions
CC associated with excess eosinophil production, the improvement
CC comprises administering a neutralising MAB for hIL-5 which does not
CC block binding of hIL-5 to the alpha-chain of the hIL-5 receptor.

CC Also claimed are methods of detecting the absence/presence of
CC hIL-5 soluble receptor alpha-chain (Ralpha)/hIL-5 complexes,
CC and of screening for compounds that antagonise the binding of
CC Ralpha/hIL-5 or Ralpha/hIL-5/MAB complex with the hIL-5 receptor
CC beta-chain. Also disclosed are novel neutralising antibodies,
CC especially altered, chimeric or humanised antibodies (see
CC W42465-66, W42471-72), as well as vectors and transformed cells
CC for their expression. These products are useful for treating
CC hIL-5-mediated disorders such as asthma, allergic rhinitis and
CC atopic dermatitis, and can also be used in the diagnosis of such
CC conditions by measurement (e.g. by ELISA) of endogenous hIL-5
CC levels.
SQ Sequence 105 AA;

Query Match 69.8%; Score 44; DB 29; Length 105;
Best Local Similarity 50.0%; Pred. No. 3.26e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 87 qndhsfpf 94
: |||||
QY 412 EOEHSFAF 419

RESULT 12
ID R29716 standard; Protein; 232 AA.
AC R29716:
DT 30-APR-1993 (first entry)
DE Guinea pig MBP-1.
KW Major basic protein; anti-asthma.
OS Cavia porcellus.
FH Key Location/Qualifiers
FT Peptide 1..15
FT /note= "signal peptide"
FT Peptide 16..113
FT /note= "prepeptide"
FT Protein 114..232
FT /note= "mature MBP-1"
PN J04320686-A.
PD 11-NOV-1992.
PF 22-APR-1991; 090704.
PR 22-APR-1991; JP-090704.
PA (SAKA) OHSUKA PHARM CO LTD.
DR WPI; 92-426680/52.
DR N-PSDB; Q31988.

PT A guinea pig MBP gene - contains DNA base sequence coding MBP
PT gene having 1 of 2 specified aminoacid sequences
PS Disclosure; Page 15; 20pp; Japanese.
CC The sequence shows the protein sequence of the guinea pig major basic
CC protein (MBP)-1. The sequence was obt'd. by first isolating MBP-1
CC and MBP-2 from the abdominal transudate of male guinea pigs treated
CC with polymyxin and ascaris extract. MBP-1 and MBP-2 were used to
CC design probes for screening of a guinea pig eosinophil cDNA library.
CC The cDNA encoding the MBPs was used to transform suitable host
CC cells which may then be cultured to obtain large amts. of MBP.
CC The MBPs may be used for the confirmation of the effect of an anti-
CC asthma drug. See also R29717.
SQ Sequence 232 AA;

Query Match 69.8%; Score 44; DB 6; Length 232;
Best Local Similarity 100.0%; Pred. No. 3.26e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 145 hsfafn 150
: |||||
QY 415 HSFAPN 420

RESULT 13
ID W40075 standard; Protein; 233 AA.
AC W40075:
DT 29-MAR-1998 (first entry)
DE Guinea pig eosinophil granule MBP GMBP-1.
KW Eosinophil-derived basic protein; EBPB; diagnosis; prevention; treatment;

KW disease; eosinophil accumulation; granule release; allergic reaction;
KW inflammatory reaction; eosinophilias; parasitic infection; pregnancy;
KW cytolytic agent; bacterial infection; cancer; major basic protein;
KW MBP; guinea pig.
OS Cavia sp.
PN US5728820-A.
PD 17-MAR-1998.
PF 23-OCT-1996; 740036.
PR 23-OCT-1996; US-740036.
PA (INCY-) INCYTE PHARM INC.
PI Akcrblom IE;
DR WPI; 98-205622/18.
PT New DNA encoding eosinophil-derived basic protein - useful for
PT producing recombinant protein used for treating, e.g. parasitic
PT infections
PS Disclosure: Fig 2: 26pp; English.
CC This sequence represents the guinea pig eosinophil granule major basic
CC protein (MBP), GMBP-1, which is used to identify an eosinophil-derived
CC basic protein (EBPH) derived from interleukin-5 (IL-5) cultured human
CC umbilical cord blood cells. This protein can be used in the diagnosis,
CC prevention and treatment of diseases and conditions associated with
CC eosinophil accumulation and granule release, including late phase
CC allergic/inflammatory reactions, eosinophilias, parasitic infections and
CC conditions associated with placental-derived, eosinophil-derived basic
CC protein accumulation in pregnancy. The protein can also be used for
CC producing recombinant EBPH, which may be useful as a cytolytic agent in
CC the treatment of e.g. bacterial and parasitic infections and cancer.
SQ Sequence 233 AA;

Query Match 69.8%; Score 44; DB 29; Length 233;
Best Local Similarity 100.0%; Pred. No. 3.26e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 146 hsfafn 151
| | | | |
Qy 415 HSFAPN 420

RESULT 14
ID P82495 standard; protein; 17 AA.
AC P82495;
DT 01-NOV-1990 (first entry)
DE Cyclic peptide with ANF activity; example 16.
KW Atrial natriuretic factor; ANF; hypertension; vasodilation.
OS Synthetic.
FH Key Location/Qualifiers
FT modified_site 1..1
FT /label-OTHER
FT /note="acetamidocysteine"
FT modified_site 6..6
FT /label-OTHER
FT /note="methionine sulphone"
FT modified_site 7..7
FT /labelaib
PN EP-292256-A.
PD 23-NOV-1988.
PF 18-MAY-1988; 304492.
PR 19-MAY-1987; US-51981.
PA (MERI) Merck and Co Inc.
PI Nutt RF, Ciccarone TM, Brady SF, Veber DF;
DR WPI; 88-331800/47.
PT New cyclic peptides with atrial natriuretic factor-like activity - useful
PT in treatment of hypertension, congestive heart failure, and disorders of
PT electrolyte imbalance and altered vascular resistance.
PS Disclosure: p; English.
CC The peptide is one of 23 disclosed, all of which are analogues of
CC the 17-membered ring portion of ANF with the cysteine moiety replaced
CC with dipeptidyl moieties or with alpha-aminoheptanoic acid. All the
CC analogues have increased potencies and metabolic stability.
CC AA3(Ala), AA5(Arg), AA12(Glu), and AA16(Ala) are D-isomers.
CC See also P82490-P82499.
SQ Sequence 17 AA;

Query Match 68.3%; Score 43; DB 1; Length 17;
Best Local Similarity 83.3%; Pred. No. 4.10e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

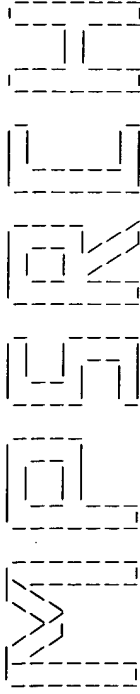
Db 12 ehafaf 17
| | | | |
Qy 414 EHSFAP 419

RESULT 15
ID W52380 standard; peptide; 16 AA.
AC W52380;
DT 01-OCT-1998 (first entry)
DE Antithrombotic peptide SEQ ID NO:26 from WO9825970.
KW Antithrombotic peptide; fibronectin type II repeat; platelet;
KW divalent cation-independent adhesion; vascular injury.
OS Homo sapiens.
OS Synthetic.
PN WO9825970-A1.
PD 18-JUN-1998.
PF 10-DEC-1997; U22724.
PR 10-DEC-1996; US-032542.
PA (UNIW) UNIV WASHINGTON.
PI Santoro SA, Staatz WD;
DR WPI; 98-348460/30.
PT New anti-thrombotic peptide(s) - used to aid in initiation of
PT recovery from vascular injury.
PS Example: Fig 6; 55pp; English.
CC The present sequence represents a peptide from the example of the
CC present invention, which describes antithrombotic peptides derived
CC from human fibronectin type II repeats. The peptides can mediate
CC divalent cation-independent adhesion of platelets to fibronectin.
CC They can therefore be used to aid in the initiation of recovery from
CC vascular injury.
SQ Sequence 16 AA;

Query Match 66.7%; Score 42; DB 33; Length 16;
Best Local Similarity 33.3%; Pred. No. 5.13e+02;
Matches 3; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 8 eqdqkysfs 16
| | | | |
Qy 412 EQHSFAPN 420

Search completed: Thu Jul 8 18:54:41 1999
Job time : 18 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:53:53 1999; MasPar time 5.19 Seconds
Tabular output not generated. 69.506 Million cell updates/sec

Title: >US-09-041-236-2
Description: (412-420) from US09041236.pep (23 of 45)
Perfect Score: 63

Sequence: 1 EQEHSFAFN 9

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.241; Variance 29.106; scale 0.764

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	51	81.0	509	2	JC5651 N-acetylglucosamine-s	3.25e+00
2	49	77.8	144	2	S73888 histidine triad prote	8.10e+00
3	49	77.8	221	2	S59832 hypothetical protein	8.10e+00
4	49	77.8	225	2	T01175 hypothetical protein	8.10e+00
5	48	76.2	280	1	C25405 ubiquinol--cytochrome	1.27e+01
6	47	74.6	780	2	A34102 von Willebrand factor	1.97e+01
7	47	74.6	796	1	JV0107 glucose dehydrogenase	1.97e+01
8	47	74.6	1407	2	B42239 adenylate cyclase (EC	1.97e+01
9	46	73.0	391	2	C71302 probable exonuclease	3.05e+01
10	46	73.0	419	1	S29127 carboxypeptidase A (E	3.05e+01
11	46	73.0	1004	2	T00046 surface layer protein	3.05e+01
12	46	73.0	1471	1	S30790 myosin MYO4 - yeast (3.05e+01
13	45	71.4	91	2	D14027 drp90 protein - soybe	4.68e+01
14	45	71.4	94	2	S71498 probable yops/flig tr	4.68e+01
15	45	71.4	183	2	S73165 H+-transporting ATP s	4.68e+01
16	45	71.4	370	1	A48252 neurotrophin allatost	4.68e+01
17	45	71.4	1048	1	YGVCAR angR protein - Vibrio	4.68e+01
18	45	71.4	1232	2	B70556 probable respiratory	4.68e+01
19	44	69.8	233	1	S13625 eosinophil major basi	7.13e+01
20	44	69.8	536	2	S62110 lysophospholipase (EC	7.13e+01
21	44	69.8	582	2	S73764 hypothetical protein	7.13e+01
22	44	69.8	619	2	G71861 prinosomal protein n'	7.13e+01
23	44	69.8	619	2	C64568 prinosomal protein re	7.13e+01

24	44	69.8	637	2	C71712 cell division protein	7.13e+01
25	44	69.8	640	2	S48423 hypothetical protein	7.13e+01
26	44	69.8	725	2	S64124 probable membrane pro	7.13e+01
27	44	69.8	754	2	S48020 kinesin-related prote	7.13e+01
28	44	69.8	1306	2	S22624 aggregation protein a	7.13e+01
29	44	69.8	1942	2	B71426 hypothetical protein	7.13e+01
30	43	68.3	240	2	S51744 testis determining fa	1.08e+02
31	43	68.3	240	2	S51713 sex determining regio	1.08e+02
32	43	68.3	394	2	D70550 probable fadE19 prote	1.08e+02
33	43	68.3	410	1	FWPMBV vicillin B precursor	1.08e+02
34	43	68.3	438	2	S35757 vicillin, 47K - garden	1.08e+02
35	43	68.3	449	2	A70568 probable dicarboxylat	1.08e+02
36	43	68.3	471	2	S45068 53K glycoprotein - ch	1.08e+02
37	43	68.3	472	2	A33312 sarcoplasmic reticulu	1.08e+02
38	43	68.3	539	2	D71522 hypothetical protein	1.08e+02
39	43	68.3	628	2	G64760 prPE protein - Escher	1.08e+02
40	43	68.3	702	2	A46075 1,4-alpha-glucan bran	1.08e+02
41	43	68.3	799	2	S56556 ribonucleoside-diphos	1.08e+02
42	43	68.3	908	2	A33280 sarcalumenin precurs	1.08e+02
43	43	68.3	1062	2	S46367 protein kinase CDC7 (1.08e+02
44	43	68.3	1101	2	S58108 hypothetical protein	1.08e+02
45	43	68.3	2156	1	RRVUNE genome polyprotein -	1.08e+02

ALIGNMENTS

RESULT 1
ENTRY JC5651 #type complete
TITLE N-acetylglucosamine-specific transporter II Nag - Vibrio
ALTERNATE_NAMES NAgE protein
ORGANISM #formal_name Vibrio cholerae non-O1
DATE 28-Oct-1997 #sequence_revision 28-Oct-1997 #text_change
17-Mar-1999
ACCESSIONS JC5651
REFERENCE JC5649
#authors Yamano, N.; Oura, N.; Wang, J.; Fujishima, S.
#journal Biosci. Biotechnol. Biochem. (1997) 61:1349-1353
#title Cloning and sequencing of the genes for N-acetylglucosamine
use that construct divergent operons (nagE-nagAC) from
Vibrio cholerae non-O1.
#cross-references MUID:97446530
#accession JC5651
#molecule_type DNA
##residues 1-509 #label YAM
##cross-references DBJ:D87820; NID:g2541899; PID:d1023703; PID:g2541900
##experimental_source strain 1148A
COMMENT This protein play a role of transporting N-acetylglucosamine into the
cells by phosphoenolpyruvate transport system.
GENETICS
#gene nagE
SUMMARY #length 509 #molecular-weight 53586 #checksum 4452
Query Match 81.0%; Score 51; DB 2; Length 509;
Best Local Similarity 75.0%; Pred. No. 3.25e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 225 QEHFAFK 232
|||:||||
Qy 413 QEHSAFN 420
RESULT 2
ENTRY S73888 #type complete
TITLE histidine triad protein A65_orf144 - Mycoplasma pneumoniae
ALTERNATE_NAMES (ATCC 29342) (Scc3)
ORGANISM #formal_name Mycoplasma pneumoniae
#variety ATCC 29342
DATE 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change
17-Jul-1998
ACCESSIONS S73888
REFERENCE S73327

```
#authors      Himmelmreich, R.; Hilbert, H.; Plagens, H.; Pirkil, E.; Li,
#journal      B.C.; Herrmann, R.
#title        Nucleic Acids Res. (1996) 24:4420-4449
#description  Complete sequence analysis of the genome of the bacterium
#organism     Mycoplasma pneumoniae.
#cross-references MUID:97105885
#accession    S73888
#status       preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
#residues     1-144 #label HITM
#cross-references EMBL:AE000055; GB:U00089; NID:g1674254; PID:g1674261
#note         the nucleotide sequence was submitted to the EMBL Data
              Library, November 1996
GENETICS
#gene         hit1
#genetic_code SGC3
CLASSIFICATION #superfamily protein kinase C inhibitor; histidine triad
              homology
FEATURE
12-112       #domain histidine triad homology #label HITN
105-109      #region histidine triad motif
SUMMARY      #length 144 #molecular-weight 16086 #checksum 5676
Query Match  77.8%; Score 49; DB 2; Length 144;
Best Local Similarity 75.0%; Pred. No. 8.10e+00;
Matches      6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 29 ENEHAFAF 36
Qy 412 QEHSFAF 419

RESULT 3
ENTRY
TITLE  S59832 #type complete
       hypothetical protein YPR174c - yeast (Saccharomyces
       cerevisiae)
ALTERNATE_NAMES
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     13-Jun-1996 #sequence_revision 01-Mar-1996 #text_change
12-Dec-1997
ACCESSIONS S59832
REFERENCE  S59829
#authors   Pauley, A.
#submission submitted to the EMBL Data Library, April 1995
#description The sequence of S. cerevisiae cosmid 9705.
#accession  S59832
#molecule_type DNA
#residues  1-221 #label PAU
#cross-references EMBL:U25842; NID:g786312; PID:g786323; MIPS:YPR174c
GENETICS
#map_position 16R
SUMMARY      #length 221 #molecular-weight 25411 #checksum 1862
Query Match  77.8%; Score 49; DB 2; Length 221;
Best Local Similarity 62.5%; Pred. No. 8.10e+00;
Matches      5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 47 QOAFGFN 54
Qy 413 QEHSFAFN 420

RESULT 4
ENTRY
TITLE  T01175 #type complete
       hypothetical protein T26D22.2 - Arabidopsis thaliana
       cress
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
DATE     12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
12-Feb-1999
ACCESSIONS T01175
REFERENCE  214253
#authors   Becker, M.; Hinds, K.
```

```
#submission   submitted to the EMBL Data Library, April 1998
#description  The sequence of A. thaliana T26D22.
#accession    T01175
#status       preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues     1-225 #label BEC
#cross-references EMBL:AF058826; NID:g3047088; PID:g3047092
GENETICS
#map_position IV
#introns      36/3; 100/1
#note         T26D22.2
SUMMARY      #length 225 #molecular-weight 25992 #checksum 3672
Query Match  77.8%; Score 49; DB 2; Length 225;
Best Local Similarity 62.5%; Pred. No. 8.10e+00;
Matches      5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 188 EOHTYGF 195
Qy 412 QEHSFAF 419

RESULT 5
ENTRY
TITLE  C25405 #type complete
       ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome c1
       precursor - Rhodobacter capsulatus (strain GA)
ORGANISM     #formal_name Rhodobacter capsulatus
DATE         05-Oct-1988 #sequence_revision 22-Jul-1994 #text_change
04-Oct-1996
ACCESSIONS   C25405; E25405
REFERENCE    A91162
#authors     Gabellini, N.; Seibald, W.
#journal     Eur. J. Biochem. (1986) 154:569-579
#title       Nucleotide sequence and transcription of the fbc operon from
              Rhodospseudomonas sphaeroides. Evaluation of the deduced
              amino acid sequences of the fbc protein, cytochrome b and
              cytochrome c-1.
#cross-references MUID:86136096
#note        source is designated as Rhodospseudomonas sphaeroides
#accession   C25405
#molecule_type DNA
#residues    1-280 #label GAB1
#accession   E25405
#molecule_type protein
#residues    22-39 #label GAB2
GENETICS
#gene        fbcC; petC
CLASSIFICATION #superfamily cytochrome c1 heme protein; cytochrome c1 heme
              protein homology
KEYWORDS      chromoprotein; electron transfer; heme; iron; oxidoreductase;
              respiratory chain; transmembrane protein
FEATURE
1-21         #domain signal sequence #status predicted #label SIG\
22-280       #product cytochrome c1 #status experimental #label MAT\
24-276       #domain cytochrome c1 heme protein homology #label C1H\
249-266      #domain transmembrane #status predicted #label TMW\
55-58        #binding site heme (Cys) (covalent) #status predicted\
59,205       #binding_site heme iron (His, Met) (axial ligands)
              #status predicted
SUMMARY      #length 280 #molecular-weight 30382 #checksum 7129
Query Match  76.2%; Score 48; DB 1; Length 280;
Best Local Similarity 50.0%; Pred. No. 1.27e+01;
Matches      4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 26 QDHAFSE 33
Qy 413 QEHSFAFN 420

RESULT 6
ENTRY
TITLE  A34102 #type fragments
       von Willebrand factor - bovine (fragments)
```



```

ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Mar-1990 #sequence_revision 05-May-1995 #text_change
30-Jan-1998

ACCESSIONS A34102; S14768; A39540; S27196; S36600; S68346
REFERENCE A34102
#authors Takagi, J.; Kasahara, K.; Sekiya, F.; Inada, Y.; Saito, Y.
#journal J. Biol. Chem. (1989) 264:10425-10430
#title A collagen-binding glycoprotein from bovine platelets is
identical to propolypeptide of von Willebrand factor.
#cross-references MUID:89278101
#accession A34102
#status preliminary
#molecule_type protein
#residues 1-21:78-98 #label TAK

REFERENCE S14768
#authors Fujisawa, T.; Takagi, J.; Sekiya, F.; Goto, A.; Mlake, F.;
Saito, Y.
#journal Eur. J. Biochem. (1991) 196:673-677
#title Monoclonal antibodies that inhibit binding of propolypeptide
of von Willebrand factor to collagen. Localization of
epitopes.
#cross-references MUID:91192039
#accession S14768
#status preliminary
#molecule_type protein
#residues 22-51:52-77,112-149 #label FUJ

REFERENCE A39540
#authors Takagi, J.; Fujisawa, T.; Sekiya, F.; Saito, Y.
#journal J. Biol. Chem. (1991) 266:5575-5579
#title Collagen-binding domain within bovine propolypeptide of von
Willebrand factor.
#cross-references MUID:91170224
#accession A39540
#status preliminary
#molecule_type protein
#residues 99-257:258-264 #label TA2

REFERENCE S27196
#authors Bakhshi, M.R.; Myers, J.C.; Howard, P.S.; Soprano, D.R.;
Kirby, E.P.
#journal Biochim. Biophys. Acta (1992) 1132:325-328
#title Sequencing of the primary adhesion domain of bovine von
Willebrand factor.
#cross-references MUID:93041938
#accession S27196
#molecule_type mRNA
#residues 265-780 #label BAK
#cross-references EMBL:X63820

REFERENCE S36600
#authors Bakhshi, M.R.
#submission submitted to the EMBL Data Library, January 1992
#accession S36600
#molecule_type mRNA
#residues 265-615,'SRVA',620-687,'M',689-718,'G',720-780 #label
BA2
#cross-references EMBL:X63820

REFERENCE S68346
#authors Takagi, J.; Aoyama, T.; Ueki, S.; Ohba, H.; Saito, Y.;
Lorand, L.
#journal Eur. J. Biochem. (1995) 232:773-777
#title Identification of factor-XIIa-reactive glutamyl residues
in the propolypeptide of bovine von Willebrand factor.
#cross-references MUID:96028118
#accession S68346
#molecule_type protein
#residues 22-33;'LSPVAVGRTCGLGNYN',99-101,'X',103-122 #label TAI
#CLASSIFICATION #superfamily von Willebrand factor; von Willebrand factor
type A repeat homology; von Willebrand factor type C repeat
homology
KEYWORDS blood coagulation; cell binding; connective tissue;
extracellular matrix; glycoprotein; plasma; von Willebrand
disease
FEATURE 513-681
#domain von Willebrand factor type A repeat homology

```

```

755 #label VWAL\
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 780 #checksum 5235

Query Match 74.6% Score 47; DB 2; Length 780;
Best Local Similarity 71.4%; Pred. No. 1.97e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 73 QDHSFDF 79
QY 413 QEHSFAF 419
I:|I|I|I|

RESULT 7
ENTRY JVO107
TITLE glucose dehydrogenase (pyrroloquinoline-quinone) (EC
1.1.99.17) - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 30-Jun-1992 #sequence_revision 12-Dec-1997 #text_change
22-May-1998
ACCESSIONS D64735; JVO107; A45997; S45201; I41228
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession D64735
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-796 #label BLAT
#cross-references GB:AE000122; GB:U00096; NID:g1786315; PID:g1786316;
#experimental_source strain K-12, substrain MG1655

REFERENCE JVO107
#authors Cleton-Jansen, A.M.; Goosen, N.; Fayet, O.; van de Putte, P.
#journal J. Bacteriol. (1990) 172:6308-6315
#title Cloning, mapping, and sequencing of the gene encoding
Escherichia coli quinoprotein glucose dehydrogenase.
#cross-references MUID:91035240
#accession JVO107
#molecule_type DNA
#residues 1-58,'L',60-148,'H',150,'KRRCHT',157-192,'K',194-796
#cross-references GB:X51323; NID:g41553; PID:g41554
#experimental_source strain K12
#label CLE

REFERENCE A45997
#authors Yamada, M.; Sumi, K.; Matsushita, K.; Adachi, O.; Yamada, Y.
#journal J. Biol. Chem. (1993) 268:12812-12817
#title Topological analysis of quinoprotein glucose dehydrogenase in
Escherichia coli and its ubiquinone-binding site.
#cross-references MUID:93286127
#accession A45997
#status preliminary
#molecule_type protein
#residues 1-20 #label YAM

REFERENCE S45181
#authors Fujita, N.
#submission submitted to the EMBL Data Library, January 1994
#accession S45201
#molecule_type DNA
#residues 1-796 #label FUJ
#cross-references EMBL:D26562; NID:g473770; PID:d1006124; PID:g473791
#experimental_source strain K-12 substrain W3110
I41228
REFERENCE
#authors Yamada, M.; Asaoka, S.; Saiter, M.H.; Yamada, Y.
#journal J. Bacteriol. (1993) 175:568-571
#title Characterization of the gcd gene from Escherichia coli K-12
W3110 and regulation of its expression.

```

```

#cross-references MUID:93123180
#accession I41228
##status preliminary
##molecule_type DNA
##residues 1-148,'H',150,'KRRCHT',157-192,'K',194-665,'H',667-796
##cross-references GB:DI2651; NID:g216555; PID:g216556
GENETICS
#gene gcd
#map_position 3 min
FUNCTION
#description catalyzes the oxidation of D-glucose to D-gluconic acid by
ubiquinone
#pathway respiratory chain
CLASSIFICATION
#superfamily glucose dehydrogenase (pyrroloquinoline-quinone)
KEYWORDS
oxidoreductase; pyrroloquinoline quinone; respiratory chain;
transmembrane protein
FEATURE
11-37 #domain transmembrane #status predicted #label TM1\
41-59 #domain transmembrane #status predicted #label TM2\
63-81 #domain transmembrane #status predicted #label TM3\
96-110 #domain transmembrane #status predicted #label TM4\
120-140 #domain transmembrane #status predicted #label TM5\
93-95 #binding_site ubiquinone (Arg, Asp) #status predicted\
466 #active_site Asp #status predicted
SUMMARY
#length 796 #molecular-weight 86747 #checksum 9982
Query Match 74.6%; Score 47; DB 1; Length 796;
Best Local Similarity 62.5%; Pred. No. 1.97e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 392 DEHTFTFN 399
:||||:
QY 413 QEHSAFN 420
RESULT 8
ENTRY
TITLE B42239 #type complete
adenylate cyclase (EC 4.6.1.1) aggregation stage - slime mold
(Dictyostelium discoideum)
ORGANISM
#formal_name Dictyostelium discoideum
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
B42239
ACCESSIONS
REFERENCE A42239
#authors Pitt, G.S.; Milona, N.; Borleis, J.; Lin, K.C.; Reed, R.R.;
Devreotes, P.N.
#journal Cell (1992) 69:305-315
#title Structurally distinct and stage-specific adenylyl cyclase
genes play different roles in Dictyostelium development.
#accession B42239
#status preliminary; not compared with conceptual translation
##molecule_type DNA; mRNA
##residues 1-1407 #label PIT
##cross-references GB:L05499; GB:M87279; NID:g167660; PID:g457431
CLASSIFICATION
#superfamily guanylate cyclase catalytic domain homology
KEYWORDS phosphorus-oxygen lyase; transmembrane protein
FEATURE
1132-1358 #domain guanylate cyclase catalytic domain homology
#label GCC
SUMMARY
#length 1407 #molecular-weight 159699 #checksum 6544
Query Match 74.6%; Score 47; DB 2; Length 1407;
Best Local Similarity 44.4%; Pred. No. 1.97e+01;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 699 DQHDYEFH 707
:||||:
QY 412 QEHSAFN 420
RESULT 9
ENTRY
TITLE C71302 #type complete
probable exonuclease - syphilis spirochete
#formal_name Treponema pallidum subsp. pallidum #common_name
syphilis spirochete
DATE 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
17-Mar-1999
ACCESSIONS
REFERENCE C71302
#authors Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.;
Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton,
R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McGood,
M.P.; Salzberg, S.; Peterson, J.; Khalek, H.; Richardson,
D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald,
L.; Artach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.;
Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey,
L.; Weidman, J.; Smith, H.O.; Venter, J.C.
#journal Science (1998) 281:375-388
#title Complete genome sequence of Treponema pallidum, the syphilis
spirochete.
#cross-references MUID:98332770
#accession C71302
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-391 #label COL
##cross-references GB:AE001237; GB:AE000520; NID:g3322916; PID:g3322921
##experimental_source strain Nichols
GENETICS
#gene TP0626
#length 391 #molecular-weight 44297 #checksum 4520
SUMMARY
Query Match 73.0%; Score 46; DB 2; Length 391;
Best Local Similarity 62.5%; Pred. No. 3.05e-01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 328 QENAFAD 335
:||||:
QY 413 QEHSAFN 420
RESULT 10
ENTRY
TITLE S29127 #type complete
carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human
pancreatic carboxypeptidase A1
ALTERNATE_NAMES
ORGANISM #formal_name Homo sapiens #common_name man
DATE 25-Feb-1994 #sequence_revision 19-Jan-1996 #text_change
17-Apr-1998
ACCESSIONS
REFERENCE S29127; A34205; S08253; S02810; S71394; S02811
#authors Catasus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.;
Wicker-Planquart, C.; Puigserver, A.
#journal Biochem. J. (1992) 287:299-303
#title cDNA cloning and sequence analysis of human pancreatic
procarboxypeptidase A1.
#accession S29127
##molecule_type mRNA
##residues 1-419 #label CAT
##cross-references EMBL:X67318; NID:g35329; PID:g35330
REFERENCE
A34205
#authors Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.
#journal Am. J. Hum. Genet. (1990) 46:793-800
#title Human carboxypeptidase A identifies a BglII RFLP and maps to
7q31-qter.
#cross-references MUID:90196012
#accession A34205
##status preliminary; not compared with conceptual translation
##molecule_type DNA
##residues 330-396 #label STE
##note the authors translated the codon CTG for residue 391 as
Val
REFERENCE
S08253
#authors Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
#journal FEBS Lett. (1990) 261:179-183
#title Further studies on the human pancreatic binary complexes

```

GENETICS	#gene	slmA
----------	-------	------

```
09-Sep-1997
S14027; S16165
#accessions
#reference
#authors
#journal
#title
#cross-references MUID:91232918
#accession
#status
#molecule_type mRNA
#residues 1-91 ##label BER1
#cross-references EMBL:X54365
REFERENCE
#authors Bergmann, J.E.
#submission submitted to the EMBL Data Library, July 1990
#accession S16165
#molecule_type mRNA
#residues 1-47,'S',49-50,'WV',53-91 ##label BER2
#cross-references EMBL:X54365; NID:g18573; PID:g18574
GENETICS
#gene drp90
#keywords DNA binding; transcription regulation
SUMMARY
#length 91 #molecular-weight 10358 #checksum 5056
Query Match 71.4%; Score 45; DB 2; Length 91;
Best Local Similarity 100.0%; Pred. No. 4.68e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 EHSFAF 8
|||||
QY 414 EHSFAF 419

RESULT 14
ENTRY D71498 #type complete
TITLE probable yops/fliq translocation protein - Chlamydia
trachomatis (serotype D, strain UW3/Cx)
#formal_name Chlamydia trachomatis
ORGANISM 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
DATE 21-Nov-1998
ACCESSIONS D71498
REFERENCE A71570
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:754-759
#title Genome sequence of an obligate intracellular pathogen of
humans: Chlamydia trachomatis.
#cross-references MUID:99000809
#accession D71498
#status preliminary
#molecule_type DNA
#residues 1-94 ##label ARN
#cross-references GB:AE001327; GB:AE001273; NID:g3328999; PID:g3329004
#experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene yscS
SUMMARY
#length 94 #molecular-weight 10573 #checksum 7843
Query Match 71.4%; Score 45; DB 2; Length 94;
Best Local Similarity 71.4%; Pred. No. 4.68e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 51 QEQTFAP 57
|||||
QY 413 QEHSFAF 419

RESULT 15
ENTRY S73165 #type complete
TITLE H+-transporting ATP synthase (EC 3.6.1.34) chain b - red alga
```

```
(Porphyra purpurea) chloroplast
#formal_name chloroplast Porphyra purpurea
19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
31-Oct-1997
ACCESSIONS S73165
REFERENCE S73108
#authors Reith, M.; Munholland, J.
#journal Plant Mol. Biol. Rep. (1995) 13:333-335
#title Complete nucleotide sequence of the Porphyra purpurea
chloroplast genome.
#accession S73165
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-183 ##label REI
#cross-references EMBL:U38804; NID:g1276652; PID:g1276710
#note the nucleotide sequence was submitted to the EMBL Data
Library, October 1995
GENETICS
#gene atpF
#genome chloroplast
CLASSIFICATION #superfamily H+-transporting ATP synthase chain I
KEYWORDS ATP biosynthesis; chloroplast; hydrolase; membrane-associated
complex; thylakoid
SUMMARY
#length 183 #molecular-weight 20591 #checksum 4734
Query Match 71.4%; Score 45; DB 2; Length 183;
Best Local Similarity 71.4%; Pred. No. 4.68e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 20 EHTFGFN 26
|||||
QY 414 EHSFAFN 420

Search completed: Thu Jul 8 18:54:02 1999
Job time : 9 secs.
```

WIREH

(TW)

Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:52:54 1999; MasPar time 3.38 Seconds

Tabular output not generated.
75.217 Million cell updates/sec

Title: >US-09-041-236-2
Description: (412-420) from US09041236.pap (23 of 45)
Perfect Score: 63
Sequence: 1 EQHSPAFN 9

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 22.870; Variance 26.251; scale 0.871

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	49	77.8	144	1	YHIT_MYCPN HYPOTHETICAL 16.1 KD P	2.88e+00
2	47	74.6	796	1	DHG_ECOLI GLUCOSE DEHYDROGENASE	7.70e+00
3	47	74.6	1407	1	CYAA_DICDI ADENYLATE CYCLASE, AGG	7.70e+00
4	46	73.0	419	1	CBP1_HUMAN CARBOXYPEPTIDASE A1 PR	1.25e+01
5	46	73.0	1471	1	MY84_YEAST MYOSIN-4 ISOFORM	1.25e+01
6	45	71.4	90	1	DP90_SOYBN DNA-BINDING PROTEIN DR	2.00e+01
7	45	71.4	183	1	ATPF_PORPU ATP SYNTHASE B CHAIN (2.00e+01
8	45	71.4	370	1	ALLS_DIPPU ALLATOSTATINS PRECURSO	2.00e+01
9	45	71.4	1048	1	ANGR_VIBAN ANGR PROTEIN	2.00e+01
10	44	69.8	233	1	EMBI_CAVPO EOSINOPHIL GRANULE MAJ	3.18e+01
11	44	69.8	536	1	SPO1_YEAST SPOROULATION PROTEIN SP	3.18e+01
12	44	69.8	582	1	Y2B1_MYCPN HYPOTHETICAL PROTEIN M	3.18e+01
13	44	69.8	619	1	PRIA_HELPY PRIMOSOMAL PROTEIN N'	3.18e+01
14	44	69.8	640	1	YIF6_YEAST HYPOTHETICAL 71.4 KD P	3.18e+01
15	44	69.8	725	1	YGL4_YEAST HYPOTHETICAL 80.0 KD P	3.18e+01
16	44	69.8	754	1	KATC_ARATH KINSEIN-LIKE PROTEIN C	3.18e+01
17	44	69.8	780	1	MSH1_SCHPO MUTS PROTEIN HOMOLOG 1	3.18e+01
18	43	68.3	240	1	SRV_SHEEP SEX-DETERMINING REGION	5.01e+01
19	43	68.3	240	1	SRV_CAPHI SEX-DETERMINING REGION	5.01e+01
20	43	68.3	251	1	EIA_ADE41 EARLY EIA 27 KD PROTEI	5.01e+01
21	43	68.3	302	1	MCRI_YEAST NADH-CYTOCHROME B5 RED	5.01e+01
22	43	68.3	410	1	VCUB_PEA PROVINCILIN PRECURSOR (5.01e+01
23	43	68.3	463	1	VCL_VICFA VICILIN PRECURSOR.	5.01e+01

24	43	68.3	544	1	YE72_METJA HYPOTHETICAL PROTEIN M	5.01e+01
25	43	68.3	628	1	PRPE_ECOLI PRPE PROTEIN.	5.01e+01
26	43	68.3	640	1	PRPE_SALTY PRPE PROTEIN.	5.01e+01
27	43	68.3	702	1	GLGB_HUMAN 1,4-ALPHA-GLUCAN BRANC	5.01e+01
28	43	68.3	908	1	SARCALUMENIN PRECURSOR	5.01e+01
29	43	68.3	961	1	UBIQUITIN FUSION DEGRA	5.01e+01
30	43	68.3	1062	1	CC7_SCHPO CELL DIVISION CONTROL	5.01e+01
31	43	68.3	1101	1	Y4G_SCHPO HYPOTHETICAL 126.5 KD	5.01e+01
32	43	68.3	2156	1	RRPL_PUUMH RNA-DIRECTED RNA POLYM	5.01e+01
33	42	66.7	155	1	Y454_MYCGE HYPOTHETICAL PROTEIN M	7.83e+01
34	42	66.7	174	1	ILIX_BOVIN INTERLEUKIN-1 RECEPTOR	7.83e+01
35	42	66.7	182	1	YC21_YEAST HYPOTHETICAL 21.1 KD P	7.83e+01
36	42	66.7	200	1	NUPL_XENLA NUCLEOPLASMIN.	7.83e+01
37	42	66.7	251	1	YPUG_BACSU HYPOTHETICAL 29.6 KD P	7.83e+01
38	42	66.7	396	1	ARRS_RANCA S-ARRESTIN (RETINAL S-	7.83e+01
39	42	66.7	396	1	ARRS_RANPI S-ARRESTIN (RETINAL S-	7.83e+01
40	42	66.7	418	1	YMT3_YEAST HYPOTHETICAL 47.3 KD P	7.83e+01
41	42	66.7	430	1	YHT2_YEAST HYPOTHETICAL 49.8 KD P	7.83e+01
42	42	66.7	441	1	YBE6_YEAST HYPOTHETICAL 50.7 KD P	7.83e+01
43	42	66.7	557	1	PAXI_HUMAN PAXILLIN.	7.83e+01
44	42	66.7	559	1	PAXI_CHICK PAXILLIN.	7.83e+01
45	42	66.7	1085	1	IFH1_YEAST IFH1 PROTEIN (RRP3 PRO	7.83e+01

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	144 AA.
ID	YHIT_MYCPN			
AC	P75504;			
DT	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	HYPOTHETICAL 16.1 KD PROTEIN HIT-LIKE PROTEIN.			
OS	MYCOPLASMA PNEUMONIAE.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;			
OC	MYCOPLASMATACEAE; MYCOPLASMA.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 29342 / M129;			
RX	MEDLINE: 97105885.			
RA	HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,			
RA	HERMANN R.;			
RT	"Complete sequence analysis of the genome of the bacterium Mycoplasma			
RT	pneumoniae."			
RL	NUCLEIC ACIDS RES. 24:4420-4449(1996).			
CC	-1- SIMILARITY: BELONGS TO THE HIT FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AE000055; G1674261; -			
DR	PROSITE; PS00892; HIT; 1.			
DR	PFAM; PF01230; HIT; 1.			
DR	HSSP; P49773; 1KPC.			
KW	HYPOTHETICAL PROTEIN.			
SQ	SEQUENCE 144 AA; 16086 MW; 4B6007C6 CRC32;			

Query Match 77.8%; Score 49; DB 1; Length 144;
Best Local Similarity 75.0%; Pred No. 2.88e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 29 ENEHAFAF 36
| | | | |
Oy 412 EQHSEAF 419

RESULT 2 STANDARD; PRT; 796 AA.
ID DHG_ECOLI

AC P15877;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSE DEHYDROGENASE [PYRROLOQUINOLINE-QUINONE] (EC 1.1.99.17).
GN GCD.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RC MEDLINE; 91035240.
RA CLETON-JANSEN A.-M., GOOSEN N., FAYET O., VAN DE PUTTE P.;
RT "Cloning, mapping, and sequencing of the gene encoding Escherichia
RT coli quinoprotein glucose dehydrogenase.";
RL J. BACTERIOL. 172:6308-6315(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RC MEDLINE; 93123180.
RA YAMADA M., ASAKA S., SATER M.H. JR., YAMADA Y.;
RT "Characterization of the gcd gene from Escherichia coli K-12 W3110
RT and regulation of its expression.";
RL J. BACTERIOL. 175:568-571(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RC MEDLINE; 94261430.
RA FUJITA N., MORI H., YURA T., ISIHAMA A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 2.4-4.1 min (110,917-193,643 bp) region.";
RL NUCLEIC ACIDS RES. 22:1637-1639(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RC MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL SCIENCE 277:1453-1474(1997).
RN [5]
RP TOPOLOGY.
RC MEDLINE; 93286127.
RA YAMADA M., SUMI K., MATSUSHITA K., ADACHI O., YAMADA Y.;
RT "Topological analysis of quinoprotein glucose dehydrogenase in
RT Escherichia coli and its ubiquinone-binding site.";
RL J. BIOL. CHEM. 268:12812-12817(1993).
RN [6]
RP 3D-STRUCTURE MODELLING.
RC MEDLINE; 96128046.
RA COZIER G.E., ANTHONY C.;
RT "Structure of the quinoprotein glucose dehydrogenase of Escherichia
RT coli modelled on that of methanol dehydrogenase from Methylobacterium
RT extorquens.";
RL BIOCHEM. J. 312:679-685(1995).
CC -1- FUNCTION: GDH IS PROBABLY INVOLVED IN ENERGY CONSERVATION RATHER
CC THAN IN SUGAR METABOLISM.
CC -1- CATALYTIC ACTIVITY: D-GLUCOSE + ACCEPTOR = D-GLUCONO-1,5-LACTONE +
CC + REDUCED ACCEPTOR.
CC -1- COFACTOR: PQQ.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL PQQ DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; X51323; G41554; -;
DR EMBL; D12651; G216556; -;
DR EMBL; D26582; G473791; -;
DR EMBL; AE000122; G1786316; -;
DR PIR; JY0107; JY0107.
DR ECOGENE; EG10369; GCD.
DR PROSITE; PS00363; BACTERIAL_PQQ_1; 1.
DR PROSITE; PS00364; BACTERIAL_PQQ_2; 1.
DR PFAM; PF01011; Bacterial_PQQ; 7.
DR HSSP; P38539; 4AAH.
KW OXIDOREDUCTASE; PQQ; TRANSMEMBRANE; INNER MEMBRANE; PERIPLASMIC.
FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 11 37 POTENTIAL.
FT DOMAIN 38 40 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 41 58 POTENTIAL.
FT DOMAIN 59 62 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 63 81 POTENTIAL.
FT DOMAIN 82 95 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 96 110 POTENTIAL.
FT DOMAIN 111 118 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 119 141 POTENTIAL.
FT DOMAIN 142 796 PERIPLASMIC (POTENTIAL).
FT ACT_SITE 466 466 BASE (POTENTIAL).
FT CONFLICT 59 59 R -> L (IN REF. 1).
FT CONFLICT 149 156 TLSADATP -> HLKRCRT (IN REF. 1 AND 2).
FT CONFLICT 193 193 N -> K (IN REF. 1 AND 2).
FT CONFLICT 666 666 Q -> H (IN REF. 2).
SQ SEQUENCE 796 AA; 86747 MW; 5F971DF0 CRC32;
Query Match 74.6%; Score 47; DB 1; Length 796;
Best Local Similarity 62.5%; Pred. No. 7.70e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 392 DEHTFTFN 399
QY 413 QHSAFNF 420
:|:|:|:|
RESULT 3
ID CYAA_DICDI STANDARD; PRT; 1407 AA.
AC Q03100;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE ADENYLATE CYCLASE, AGGREGATION SPECIFIC (EC 4.6.1.1) (ATP
DE PYROPHOSPHATE-LYASE) (ADENYL CYCLASE).
OS ACA OR ACA.
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92233467.
RA PITT G.S., MILONA N., BORLEIS J., LIN K.C., REED R.R., DEVREOTES P.N.;
RT "Structurally distinct and stage-specific adenyl cyclase genes play
RT different roles in Dictyostelium development.";
RL CELL 69:305-315(1992).
CC -1- FUNCTION: COORDINATES CELL AGGREGATION BY SYNTHESIZING THE CAMP
CC THAT INFLUENCES DIFFERENTIATION AND MORPHOGENESIS OF CELLS
CC WITHIN A DEVELOPING MULTICELLULAR STRUCTURE.
CC -1- CATALYTIC ACTIVITY: ATP + 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -1- ENZYME REGULATION: REGULATED BY A SURFACE RECEPTOR THROUGH A
CC GUANINE NUCLEOTIDE BINDING PROTEIN. BOTH POSITIVELY AND
CC NEGATIVELY REGULATED BY EXTRACELLULAR CAMP; THIS REGULATION IS
CC PART OF THE MECHANISM THAT ESTABLISHES THE OSCILLATORY CAMP WAVES
CC DURING AGGREGATION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS TWO
CC DOMAINS WITH 6 POTENTIAL TRANSMEMBRANE REGIONS EACH.
CC -1- DEVELOPMENTAL STAGE: CELL AGGREGATION AND LATER STAGES.
CC -1- SIMILARITY: BELONGS TO CLASS-4 OF ADENYL CYCLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; L05499; G457431; -
DR EMBL; L05496; G457431; JOINED.
DR EMBL; L05497; G457431; JOINED.
DR EMBL; L05498; G457431; JOINED.
DR PIR; B42239; B42239.
DR DICTYDB; DD02024; ACRA.
DR PROSITE; PS00452; GUANYLATE_CYCLASES; 1.
DR PFAM; PF00211; guanylate_cyc; 2.
DR HSSP; P46769; IAB8.
KW LYASE; CAMP SYNTHESIS; TRANSMEMBRANE.
FT DOMAIN 1 219 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 220 240 POTENTIAL.
FT TRANSMEM 244 264 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 304 324 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 353 373 POTENTIAL.
FT DOMAIN 374 962 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 963 979 POTENTIAL.
FT TRANSMEM 992 1012 POTENTIAL.
FT TRANSMEM 1018 1038 POTENTIAL.
FT TRANSMEM 1071 1091 POTENTIAL.
FT TRANSMEM 1105 1125 POTENTIAL.
FT TRANSMEM 1378 1398 POTENTIAL.
FT DOMAIN 1399 1407 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 65 74 POLY-GLY.
FT DOMAIN 535 576 ASN-RICH.
FT DOMAIN 753 785 POLY-ASN.
FT DOMAIN 847 864 POLY-ASN.
FT DOMAIN 88 113 ASN-RICH.
FT DOMAIN 532 576 ASN-RICH.
FT DOMAIN 753 872 ASN-RICH.
SQ SEQUENCE 1407 AA; 159699 MW; FLB7939C CRC32;

Query Match 74.6%; Score 47; DB 1; Length 1407;

Best Local Similarity 44.4%; Pred. No. 7.70e+00;

Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 699 DQEHYEFH 707

|||||:

QY 412 EQEHSFAFN 420

RESULT 4

ID CBPI_HUMAN STANDARD; PRT; 419 AA.
AC P15085;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DE 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE CARBOXYPEPTIDASE A1 PRECURSOR (EC 3.4.17.1).
GN CPA1 OR CPA.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS.
RX MEDLINE; 93038569.
RA CATASUS L., VILLEGAS V., PASCUAL R., AVILES F.X.,
RA WICKER-PLANQUART C., PUIGSERVER A.;
RT "cDNA cloning and sequence analysis of human pancreatic
RT procarboxypeptidase A1."
RL BIOCHEM. J. 287:299-303(1992).
RN [2]
RP SEQUENCE OF 17-42.
RX MEDLINE; 89153096.

RA PASCUAL R., BURGOS F.J., SORIANO F., MENDEZ E., AVILES F.X.;
RT "Purification and properties of five different forms of human
RT procarboxypeptidases";
RL EUR. J. BIOCHEM. 179:609-616(1989).
CC -!- CATALYTIC ACTIVITY: PEPTIDYL-L-AMINO ACID + H(2)O - PEPTIDE +
CC L-AMINO ACID.
CC -!- SUBUNIT: MONOMER.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; X67318; G35330; -
DR PIR; S02810; S02810.
DR PIR; S29127; S29127.
DR MIM; I14850; -
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR PFAM; PF00246; Zn_carboxypept; 1.
DR HSSP; P00730; LPYT.
KW HYDROLASE; CARBOXYPEPTIDASE; ZINC; ZMOGEN; SIGNAL.
FT SIGNAL 1 16
FT PROPEP 17 110 ACTIVATION PEPTIDE.
FT CHAIN 111 419 CARBOXYPEPTIDASE A1.
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 182 182 ZINC (BY SIMILARITY).
FT METAL 306 306 ZINC (BY SIMILARITY).
FT ACT_SITE 358 358 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 380 380 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 248 271 BY SIMILARITY.
SQ SEQUENCE 419 AA; 47140 MW; E397BA9E CRC32;

Query Match 73.0%; Score 46; DB 1; Length 419;

Best Local Similarity 75.0%; Pred. No. 1.25e+01;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 100 EQQMFAF 107

|||||:

QY 412 EQEHSFAF 419

RESULT 5

ID MYS4_YEAST STANDARD; PRT; 1471 AA.
AC P32492;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE MYOSIN-4 ISOFORM.
GN MYO4 OR SHE1 OR YAL029C OR FUN22.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94334370.
RA HAARER B.K., PETZOLD A., LILLIE S.H., BROWN S.S.;
RT "Identification of MYO4, a second class V myosin gene in yeast.";
RL J. CELL SCI. 107:1055-1064(1994).
RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C / AB972;

RX MEDLINE; 95249563.

RA BUSSEY H., KABACK D.B., ZHONG W., VO D.T., CLARK M.W., FORTIN N.

RA HALL J., OUELLETTE B.F.F., KENG T., BARTON A.B., SU Y., DAVIES C.K.,

RA STORMS R.K.;

RT "The nucleotide sequence of chromosome I from Saccharomyces

RT cerevisiae.";

```
RL PROC. NATL. ACAD. SCI. U.S.A. 92:3809-3813(1995).
CC -!- FUNCTION: SEEMS TO BE RESPONSIBLE FOR THE ACCUMULATION IN DAUGHTER
CC CELLS OF ASH1. A REPRESSOR OF THE HO ENDONUCLEASE.
CC -!- SUBCELLULAR LOCATION: ACCUMULATES PREFERENTIALLY IN GROWING BUDS.
CC -!- SIMILARITY: BELONGS TO CLASS-5 MYOSINS. STRONGEST, TO YEAST MYO2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M90057; G172024; -
DR EMBL; U12980; G595536; -
DR PIR; S30790; S30790.
DR SGD; L0001224; MYO4.
DR PFAM; PF00063; myosin_head; 1.
DR PFAM; PF00612; IQ; 5.
DR HSP; P08799; 1MND.
DR MYOSIN; COILED COIL; ATP-BINDING; ACTIN-BINDING;
KW HEPTAD REPEAT PATTERN; ALKYLATION; MULTIGENE FAMILY.
FT DOMAIN 1 777 GLOBULAR HEAD-LIKE DOMAIN.
FT DOMAIN 778 922 NECK.
FT DOMAIN 938 1063 COILED COIL.
FT DOMAIN 1064 1471 CARBOXYL-TERMINAL.
FT NP_BIND 165 172 ATP (POTENTIAL).
FT MOD_RES 688 688 ALKYLATION (BY SIMILARITY).
FT MOD_RES 698 698 ALKYLATION (BY SIMILARITY).
SQ SEQUENCE 1471 AA; 169343 MW; ED442252 CRC32;

Query Match 73.0%; Score 46; DB 1; Length 1471;
Best Local Similarity 50.0%; Pred. No. 1.25e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 437 QDHFVFSF 444
Qy 412 EQEHSFAF 419

RESULT 6
ID DP90_SOYBN STANDARD; PRT; 90 AA.
AC P24274;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE DNA-BINDING PROTEIN DRP90.
GN DRP90.
OS GLYCINE MAX (SOYBEAN).
CC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
CC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
CC FABALES; FABACEAE; PAPILIONOIDEAE; GLYCINE.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-CV. PRIZE;
RX MEDLINE; 91232918.
RA BERGMANN J.E., PREDDIE E., CORTES L., BROUSSEAU R.;
RT "A protein drp90 encoded on the leftwards strand of soybean nodule
RT urate oxidase cDNA binds to a regulatory sequence in leghemoglobin C3
RT gene.";
RL NUCLEIC ACIDS RES. 19:1338-1338(1991).
CC -!- FUNCTION: POSSIBLE ROLE IN GENE REGULATION DURING SOYBEAN
CC NODULE DIFFERENTIATION.
CC -!- BINDS TO A REGULATORY SEQUENCE IN LEGHEMOGLOBIN C3 GENE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC EMBL; X54365; G18574; -
DR PIR; S14027; S14027.
DR PIR; S16165; S16165.
KW DNA-BINDING; TRANSCRIPTION REGULATION.
FT INIT_MET 0
SQ SEQUENCE 90 AA; 10280 MW; 5B7A70C6 CRC32;

Query Match 71.4%; Score 45; DB 1; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.00e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 EHSFAF 7
Qy 414 EHSFAF 419

RESULT 7
ID ATPF_PORPU STANDARD; PRT; 183 AA.
AC P51244;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ATP SYNTHASE B CHAIN (EC 3.6.1.34) (SUBUNIT I).
GN ATPF.
OS PORPHYRA PURPUREA.
OG CHLOROPLAST.
CC EUKARYOTA; RHODOPHYTA; BANGIOPHYCEAE; BANGIALES; PORPHYRA.
RN SEQUENCE FROM N.A.
RC STRAIN-AVONPORT;
RA REITH M.E., MUNHOLLAND J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT genome.";
RL PLANT MOL. BIOL. REP. 13:333-335(1995).
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE.
CC -!- SIMILARITY: TO OTHER B SUBUNITS AND ALSO TO B' SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38804; G1276710; -
DR PFAM; PF00430; ATP-synt_B; 1.
DR MENDEL; 10257; PORPU; atpf1.1.
KW HYDROGEN ION TRANSPORT; TRANSMEMBRANE; CF(0); CHLOROPLAST.
FT TRANSMEM 28 48 POTENTIAL.
SQ SEQUENCE 183 AA; 20591 MW; B1BC43BC CRC32;

Query Match 71.4%; Score 45; DB 1; Length 183;
Best Local Similarity 71.4%; Pred. No. 2.00e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 20 EHTGFN 26
Qy 414 EHSFAF 420

RESULT 8
ID ALLS_DIPPU STANDARD; PRT; 370 AA.
AC P12764; P12765; P12766; P12767; P22410;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ALLATOSTATINS PRECURSOR.
```


OS DIPTERA PUNCTATA (PACIFIC BEETLE COCKROACH).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DICTIOPTERA; BLATTARIA; BLABEROIDEA; BLABERIDAE;
 OC DIPTERA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 94022263.
 RA DONLY B.C., DING Q., TOBE S.S., BENDENA W.G.;
 RT "Molecular cloning of the gene for the allatostatatin family of
 RT neuropeptides from the cockroach *Diploptera punctata*.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 90:8807-8811(1993).
 RN [2]
 RP SEQUENCE OF 154-161; 176-188; 192-200 AND 204-213.
 RC TISSUE-BRAIN;
 RX MEDLINE; 89345592.
 RA WOODHEAD A.P., STAY B., SEIDEL S.L., KHAN M.A., TOBE S.S.;
 RT "Primary structure of four allatostatins: neuropeptide inhibitors of
 RT juvenile hormone synthesis.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 86:5997-6001(1989).
 RN [3]
 RP SEQUENCE OF 77-94 (AST2).
 RC TISSUE-BRAIN;
 RX MEDLINE; 91172818.
 RA PRATT G.E., FARNSWORTH D.E., FOK K.F., SIEGEL N.R., MCCORMACK A.L.,
 RA SHABANOWITZ J., HUNT D.F., FEYEREISEN R.;
 RT "Identity of a second type of allatostatatin from cockroach brains: an
 RT octadecapeptide amide with a tyrosine-rich address sequence.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:2412-2416(1991).
 RN [4]
 RP SEQUENCE OF 77-189 FROM N.A.
 RC MEDLINE; 95083607.
 RA REICHWALD K., UNNITHAN G.C., DAVIS N.T., AGRICOLA H.,
 RA FEYEREISEN R.;
 RT "Expression of the allatostatatin gene in endocrine cells of the
 RT cockroach midgut.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 91:11894-11898(1994).
 RN [5]
 RP SEQUENCE OF 176-188 (AST7).
 RC TISSUE-BRAIN;
 RX MEDLINE; 89392029.
 RA PRATT G.E., FARNSWORTH D.E., SIEGEL N.R., FOK K.F., FEYEREISEN R.;
 RT "Identification of an allatostatatin from adult *Diploptera punctata*.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 163:1243-1247(1989).
 CC -1- FUNCTION: NEUROPEPTIDE INHIBITORS OF JUVENILE HORMONE SYNTHESIS.
 CC -1- TISSUE SPECIFICITY: BRAIN, SUBESOPHAGEAL GANGLION AND CORPUS
 CC ALLATUM.
 CC -1- SIMILARITY: TO CALLATOSTATINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U00444; G416119; -.
 DR EMBL; S74578; G786389; -.
 DR PIR; A33259; A33259.
 DR PIR; A33550; A33550.
 DR PIR; A39163; A39163.
 DR PIR; B33550; B33550.
 DR PIR; C33550; C33550.
 DR PIR; D33550; D33550.
 KW NEUROPEPTIDE; AMIDATION; CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL.
 FT SIGNAL 1 27
 FT PEPTIDE 68 73
 FT ALLATOSTATIN 1 (AST1).
 FT PEPTIDE 77 94
 FT ALLATOSTATIN 2 (AST2) (ASB2) (ASA5).
 FT PEPTIDE 98 105
 FT ALLATOSTATIN 3 (AST3).
 FT PEPTIDE 109 117
 FT ALLATOSTATIN 4 (AST4) (ASAT7).
 FT PEPTIDE 154 161
 FT ALLATOSTATIN 5 (AST5) (ASA4).
 FT PEPTIDE 165 172
 FT ALLATOSTATIN 6 (AST6).
 FT

FT PEPTIDE 176 188
 FT ALLATOSTATIN 7 (AST7) (ASA1).
 FT PEPTIDE 192 200
 FT ALLATOSTATIN 8 (AST8) (ASA3).
 FT PEPTIDE 204 213
 FT ALLATOSTATIN 9 (AST9) (ASA2).
 FT PEPTIDE 217 232
 FT ALLATOSTATIN 10 (AST10).
 FT PEPTIDE 254 264
 FT ALLATOSTATIN 11 (AST11) (ASA6).
 FT PEPTIDE 348 353
 FT ALLATOSTATIN 12 (AST12).
 FT PEPTIDE 357 364
 FT ALLATOSTATIN 13 (AST13).
 FT MOD_RES 73 73
 FT AMIDATION (G-74 PROVIDE AMIDE GROUP).
 FT MOD_RES 94 94
 FT AMIDATION (G-95 PROVIDE AMIDE GROUP).
 FT MOD_RES 105 105
 FT AMIDATION (G-106 PROVIDE AMIDE GROUP).
 FT MOD_RES 117 117
 FT AMIDATION (G-118 PROVIDE AMIDE GROUP).
 FT MOD_RES 161 161
 FT AMIDATION (G-162 PROVIDE AMIDE GROUP).
 FT MOD_RES 172 172
 FT AMIDATION (G-173 PROVIDE AMIDE GROUP).
 FT MOD_RES 188 188
 FT AMIDATION (G-189 PROVIDE AMIDE GROUP).
 FT MOD_RES 200 200
 FT AMIDATION (G-201 PROVIDE AMIDE GROUP).
 FT MOD_RES 213 213
 FT AMIDATION (G-214 PROVIDE AMIDE GROUP).
 FT MOD_RES 232 232
 FT AMIDATION (G-233 PROVIDE AMIDE GROUP).
 FT MOD_RES 264 264
 FT AMIDATION (G-265 PROVIDE AMIDE GROUP).
 FT MOD_RES 353 353
 FT AMIDATION (G-354 PROVIDE AMIDE GROUP).
 FT MOD_RES 364 364
 FT AMIDATION (G-365 PROVIDE AMIDE GROUP).
 SQ SEQUENCE 370 AA; 41562 MW; E3F1597F CRC32;

Query Match 71.4%; Score 45; DB 1; Length 370;
 Best Local Similarity 71.4%; Pred. No. 2.00e+01;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 256 QEHRSF 262
 ||| |:
 Qy 413 QEHSAF 419

RESULT 9
 ID ANGR-VIBAN STANDARD; PRT: 1048 AA.
 AC P19828;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE ANGR PROTEIN.
 GN ANGR.
 OS VIBRIO ANGUILLARUM.
 OG PLASMID PJM1.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIONACEAE; VIBRIO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-775;
 RX MEDLINE; 90185247.
 RA FARRELL D.H., MIKESELL P., ACTIS L.A., CROSA J.H.;
 RT "A regulatory gene, *angR*, of the iron uptake system of *Vibrio*
 RT *anguillarum* similarity with phage P22 *cro* and regulation by iron.";
 RL GENE 86:45-51(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-531A;
 RX MEDLINE; 93328275.
 RA TOLMANSKY M.E., ACTIS L.A., CROSA J.H.;
 RT "A single amino acid change in *angR*, a protein encoded by *pJM1*-like
 RT virulence plasmids, results in hyperproduction of *anguibactin*.";
 RL INFECT. IMMUN. 61:3228-3233(1993).
 RN [3]
 RP POSSIBLE FUNCTION.
 RA BAIRACH A.;
 RL UNPUBLISHED OBSERVATIONS (MAY-1991).
 CC -1- FUNCTION: ACCORDING TO REF.1: A TRANS-ACTING TRANSCRIPTIONAL
 CC ACTIVATOR FOR THE IRON UPTAKE SYSTEM OF *V. ANGUILLARUM*. ALLOWS THE
 CC BACTERIUM TO GROW IN THE IRON-DEFICIENT ENVIRONMENT OF THE FISH.
 CC -1- FUNCTION: ACCORDING TO REF.3: AN ENZYME INVOLVED IN THE
 CC BIOSYNTHESIS OF ANGUIBACTIN; AN IRON-BINDING SIDEROPHORE.
 CC -1- INDUCTION: ANGR TRANSCRIPTION IS REGULATED BY IRON.
 CC -1- THE SEQUENCE SHOWN IS THAT OF STRAIN 775.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration


```
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN MG281 HOMOLOG.
OS MYCOPLASMA PNEUMONIAE.
CC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
CC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE: 97103885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL pneumoniae";
CC NUCLEIC ACIDS RES. 24:4420-4449(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000043; G1674125;
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 20 40 POTENTIAL.
SQ SEQUENCE 582 AA; 65977 MW; 5083BDDF CRC32;

Query Match 69.8%; Score 44; DB 1; Length 582;
Best Local Similarity 44.4%; Pred. No. 3.18e+01;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 334 NODNFGFN 342
QY 412 EQEHSFAFN 420
:|:|:|

RESULT 13
ID PRIA_HELPY STANDARD; PRT; 619 AA.
AC O25149;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y).
GN PRIA OR HP0387.
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
CC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
CC HELICOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE: 97394467.
RA FLETSCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKI M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;
RT "The complete genome sequenced of the gastric pathogen Helicobacter
RL pylori.";
CC NATURE 388:539-547(1997).
CC -!- FUNCTION: PRIA RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX
CC SSBNA. THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS
CC PRIB AND PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE
CC SUBSEQUENT ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN
CC FUNCTIONS AS A HELICASE WITHIN THE PRIMOSOME (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PRIA SUBFAMILY OF HELICASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000555; G2313488;
DR TIGR: HP0387;
DR PFAM: PF00271; helicase_C; 1.
KW DNA REPLICATION; DNA-BINDING; ATP-BINDING; HELICASE; PRIMOSOME;
KW ZINC-FINGER.
FT NP_BIND 132 139 ATP (POTENTIAL).
FT SITE 228 231 DEEH BOX.
FT ZN_FING 336 348 C4-TYPE (POTENTIAL).
FT ZN_FING 363 379 C4-TYPE (POTENTIAL).
SQ SEQUENCE 619 AA; 70361 MW; 264AD677 CRC32;

Query Match 69.8%; Score 44; DB 1; Length 619;
Best Local Similarity 33.3%; Pred. No. 3.18e+01;
Matches 3; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 228 DEEHDFSYK 236
QY 412 EQEHSFAFN 420
:|:|:|

RESULT 14
ID YIF6_YEAST STANDARD; PRT; 640 AA.
AC P40522;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 71.4 KD PROTEIN IN SNP1-GPPI INTERGENIC REGION.
GN YI056W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
CC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
CC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA WALSH S.V., WHITEHEAD S.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38060; E121327;
DR EMBL: Z47047; G763290;
DR PIR: S48423; S48423.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 640 AA; 71420 MW; F47227DD CRC32;

Query Match 69.8%; Score 44; DB 1; Length 640;
Best Local Similarity 33.3%; Pred. No. 3.18e+01;
Matches 3; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db 456 DREQAFTH 464
QY 412 EQEHSFAFN 420
:|:|:|

RESULT 15
```

ID YGL4_YEAST STANDARD; PRT: 725 AA.
AC P33134;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 80.0 KD PROTEIN IN SNF4-TAF60 INTERGENIC REGION.
GN YGL114W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA LAUQUIN G.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z72636; E243342; -.
DR HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
KW FT TRANSMEM 28 48 POTENTIAL.
FT TRANSMEM 134 154 POTENTIAL.
FT TRANSMEM 254 274 POTENTIAL.
FT TRANSMEM 353 373 POTENTIAL.
FT TRANSMEM 449 469 POTENTIAL.
FT TRANSMEM 472 492 POTENTIAL.
FT TRANSMEM 564 584 POTENTIAL.
FT TRANSMEM 644 664 POTENTIAL.
FT TRANSMEM 697 717 POTENTIAL.
SQ SEQUENCE 725 AA; 80009 MW; 1B0643B0 CRC32;

Query Match 69.8%; Score 44; DB 1; Length 725;
Best Local Similarity 62.5%; Pred. No. 3.18e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 127 EGGOSFTF 134
||:|:|
Qy 412 EQHSFAF 419

Search completed: Thu Jul 8 18:53:02 1999
Job time : 8 secs.

(TM)

RESULT	2	
ID	O32444	PRELIMINARY;
AC	O32444	PRT; 509 AA.
IC	O32444;	
DT	01-JAN-1998	(TREMBLREL. 05, CREATED)
DT	01-JAN-1998	(TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE	NAGE.	
GN	NAGE.	
OS	VIBRIO CHOLERAE NON-O1.	
OC	BACTERIA; PROTEOBACTERIA;	GAMMA SUBDIVISION; VIBRIONACEAE; VIBRIO.
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=1148A;	
RX	MEDLINE: 97446530	

RA YAMANO N., OURA N., WANG J., FUJISHIMA S.;
 RT "Cloning and sequencing of the genes for N-acetylglucosamine use that
 RT construct divergent operons (nagE-nagAC) from *Vibrio cholerae*
 RT non-O1.";
 RL BIOSCI. BIOTECHNOL. BIOCHEM. 61:1349-1353(1997).
 DR EMBL: D87820; D1023703; -.
 DR PROSITE; PS01035; PTS_EIIB_CYS: 1.
 DR PFAM; PF00367; PTS_EIIB: 1.
 SQ SEQUENCE 509 AA; 53586 MW; 7EB6EF7F CRC32;

Query Match 81.0%; Score 51; DB 2; Length 509;
 Best Local Similarity 75.0%; Pred. No. 1.79e+00;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 225 QEHTEFAFK 232
 QY 413 QEHSAFN 420
 |||:||||

RESULT 3

ID O82183 PRELIMINARY; PRT; 242 AA.
 AC O82183;
 DT 01-NOV-1998 (TREMREL. 08, CREATED)
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
 DE T4C15.25 PROTEIN.
 GN T4C15.25.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA ROUNSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,
 RA SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;
 RT "Arabidopsis thaliana chromosome II BAC T4C15 genomic sequence.";
 RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AC004657; G3668095; -.
 SQ SEQUENCE 242 AA; 28117 MW; 6002C4E0 CRC32;

Query Match 79.4%; Score 50; DB 10; Length 242;
 Best Local Similarity 62.5%; Pred. No. 2.97e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 47 QDHTFAFS 54
 QY 413 QEHSAFN 420
 |.:|:|:

RESULT 4

ID Q06616 PRELIMINARY; PRT; 221 AA.
 AC Q06616;
 DT 01-NOV-1996 (TREMREL. 01, CREATED)
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
 DE SIMILARITY TO A NUCLEAR LAMIN FROM *C. ELEGANS*.
 GN P9705.11.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCHAROMYCETALES;
 OC SACCHAROMYCETACEAE; SACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RX MEDLINE; 97313271.
 RA BUSSEY H., STORMS R.K., AHMED A., ALBERMANN K., ALLEN E., ANSORGE W.,
 RA ARAUJO R., APARICIO A., BARRELL B., BADCOCK K., BENES V., BOTSTEIN D.,
 RA BOWMAN S., CRUCKER M., CARPENTER J., CHERRY J.M., CHUNG E.,
 RA CHURCHER C., COSTER F., DAVIS K., DAVIS R.W., DIETRICH F.S.,
 RA DELIUS H., DIPALO T., DUBOIS E., DUSTERHOFT A., DUNCAN M., FLOETH M.,
 RA FORTIN N., FRIESEN J.D., FRITZ C., GOFFEAU A., HALL J., HEBLING U.,
 RA HEUMANN K., HILBERT H., HILLIER L., HUNICKE-SMITH S., HYMAN R.,
 RA JOHNSTON M., KALMAN S., KLEINE K., KOMP C., KURDI O., LASHKARI D.,

RA LEW H., LIN A., LIN D., LOUIS E.J., MARATHE R., MESSENGUY F.,
 RA MEMES H.W., MIRIPATI S., MOSTL D., MULLER-AUER S., NAMATH A.,
 RA NEWITCH U., OEFNER P., PEARSON D., PETEL F.X., POHL T.M.,
 RA PURNELLE D., SCHAFER M., SCHAEFE M., SCHERENS B., SCHRAMM S.,
 RA SCHROEDER M., SDICU A.M., TETTELIN H., URRESTARAZU L.A., USHINSKY S.,
 RA VIERENDELS F., VISSERS S., VOSS H., WALSH S.V., WAMBUTT R., WANG Y.,
 RA WEDLER E., WEDLER H., WINNETT E., ZHONG W.W., ZOLLNER A., VO D.H.,
 RA HANI J.;
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome
 RT XVI.";
 RL NATURE 387:0-0(0).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
 RA FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J., KUCABA T.,
 RA HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D., MENEZES S.,
 RA JOHNSTON L., LANGSTON Y., LATREILLE P., LE T., MARDIS E., MENEZES S.,
 RA MILLER N., NHAN M., PAULEY A., PELUSO D., RIFKEN L., RILES L.,
 RA TAICH A., TREVASKIS E., VIGNATI D., WILCOX L., WOLDMAN P., VAUDIN M.,
 RA WILSON R., WATERSTON R.;
 RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA PAULEY A.;
 RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA WATERSTON R.;
 RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA JIA Y., CHERRY J.M.;
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U25842; G786323; -.
 SQ SEQUENCE 221 AA; 25411 MW; FDA9B3E9 CRC32;

Query Match 77.8%; Score 49; DB 3; Length 221;
 Best Local Similarity 62.5%; Pred. No. 4.88e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 47 QOHAFGEN 54
 QY 413 QEHSAFN 420
 |.:|:|:

RESULT 5

ID O65233 PRELIMINARY; PRT; 225 AA.
 AC O65233;
 DT 01-AUG-1998 (TREMREL. 07, CREATED)
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
 DE T26D22.2 PROTEIN.
 GN T26D22.2.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMERYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA WASHU;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA BECKER M., HINDS K.;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.

```
RC STRAIN-CV. COLUMBIA;
RA WATERSTON R.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF058826; G3047032; -.
SQ SEQUENCE 235 AA; 25992 MW; 6093CD58 CRC32;

Query Match 77.8%; Score 49; DB 10; Length 225;
Best Local Similarity 62.5%; Pred. No. 4.88e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 188 EOEHTYGF 195
Qy 412 EOEHSFAF 419

RESULT 6
ID Q50858 PRELIMINARY; PRT; 2605 AA.
AC Q50858;
DT 01-NOV-1996 (TREMUREL. 01, CREATED)
DT 01-NOV-1996 (TREMUREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMUREL. 08, LAST ANNOTATION UPDATE)
DE SAFRAMYCIN M1 SYNTHETASE A.
GN SAFA.
OS MYXOCOCCUS XANTHUS.
OC BACTERIA; PROTEOBACTERIA; DELTA SUBDIVISION; MYXOBACTERIA;
OC MYXOCOCCALES; CISTOBACTERINEAE; MYXOCOCCACEAE; MYXOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DM504-15;
RX MEDLINE; 96032006.
RA POSPIECH A., CLUZEL B., BIETENHADER J., SCHUPP T.;
RT "A new Myxococcus xanthus gene cluster for the biosynthesis of the
RT antibiotic saframycin M1 encoding a peptide synthetase.";
RL MICROBIOLOGY 141:1793-1803(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-DM504-15;
RX MEDLINE; 97090395.
RA POSPIECH A., BIETENHADER J., SCHUPP T.;
RT "Two multifunctional peptide synthetases and an O-methyltransferase
RT are involved in the biosynthesis of the DNA-binding antibiotic and
RT antitumour agent saframycin M1 from Myxococcus xanthus.";
RL MICROBIOLOGY 142:741-746(1996).
DR EMBL; U24657; G1171129; -.
DR PROSITE; PS00455; AMP_BINDING; 2.
DR PFAM; PF00501; AMP-binding; 2.
DR PFAM; PF00550; pp-binding; 2.
DR PFAM; PF00668; DUF4; 2.
SQ SEQUENCE 2605 AA; 285745 MW; 09C34B84 CRC32;

Query Match 76.2%; Score 48; DB 2; Length 2605;
Best Local Similarity 62.5%; Pred. No. 7.99e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1786 QNHSEFSD 1793
Qy 413 QEHSFAFN 420

RESULT 7
ID Q84464 PRELIMINARY; PRT; 256 AA.
AC Q84464;
DT 01-NOV-1996 (TREMUREL. 01, CREATED)
DT 01-NOV-1996 (TREMUREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMUREL. 08, LAST ANNOTATION UPDATE)
DE GENOME, PARTIAL SEQUENCE.
GN A144L.
OS PARAMESCIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 95407089.
RX LI Y., LU Z., BURBANK D.E., KUTISH G.F., ROCK D.L., ETEN J.L.;

RT "Analysis of 43 kb of the Chlorella virus PBCV-1 330-kb genome: map
RT positions 45 to 88.";
RL VIROLOGY 212:134-150(1995).
DR EMBL; U42580; G1131488; -.
SQ SEQUENCE 256 AA; 29097 MW; 74115113 CRC32;

Query Match 74.6%; Score 47; DB 14; Length 256;
Best Local Similarity 50.0%; Pred. No. 1.30e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 3 DNDHSFSF 10
Qy 412 EOEHSFAF 419

RESULT 8
ID O76289 PRELIMINARY; PRT; 262 AA.
AC O76289;
DT 01-NOV-1998 (TREMUREL. 08, CREATED)
DT 01-NOV-1998 (TREMUREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMUREL. 08, LAST ANNOTATION UPDATE)
DE SECRETED LECTIN HOMOLOG PRECURSOR.
GN HEEL-1.
OS HELIOLIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOLIDARIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98298333.
RA HAAG E.S., RAFF R.A.;
RT "Isolation and characterization of three mRNAs enriched in embryos of
RT the direct-developing sea urchin helioidaris erythrogramma:
RT evolution of larval ectoderm.";
RL DEV. GENES EVOL. 208:188-204(1998).
DR EMBL; AF049335; G3378108; -.
KW SIGNAL; LECTIN.
FT SIGNAL 4 18 POTENTIAL.
SQ SEQUENCE 262 AA; 29529 MW; A5FB8D99 CRC32;

Query Match 74.6%; Score 47; DB 5; Length 262;
Best Local Similarity 55.6%; Pred. No. 1.30e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 157 QOENSFAYD 165
Qy 412 EOEHSFAFN 420

RESULT 9
ID O57527 PRELIMINARY; PRT; 295 AA.
AC O57527;
DT 01-JUN-1998 (TREMUREL. 06, CREATED)
DT 01-JUN-1998 (TREMUREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMUREL. 06, LAST ANNOTATION UPDATE)
DE GASTRIC H,K-ATPASE BETA SUBUNIT.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPOIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RA CHEN P.-X., MATHEWS P.M., GOOD P.J., ROSSIER B.C., GEERING K.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF042812; G2801823; -.
SQ SEQUENCE 295 AA; 33410 MW; F6C540CF CRC32;

Query Match 73.0%; Score 46; DB 13; Length 295;
Best Local Similarity 33.3%; Pred. No. 2.09e+01;
Matches 3; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db 163 EHDHTFGYK 171
Qy 412 EOEHSFAFN 420
```

```

RESULT 10
ID O23199 PRELIMINARY; PRT: 326 AA.
AC O23199;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 36.9 KD PROTEIN.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUPHYLLOPHYTES: VIRIDIPANTAE: STREPTOPHYTA: TRACHEOPHYTA:
OC EUPHYLLOPHYTES: SPERMATOPHYTA: MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
[1]
RN SEQUENCE FROM N.A.
RA BEVAN M., TERRY N., VOS P., HEIJNEN L., MEWES H.W., SCHUELLER C.,
RA CHALWATZIS N.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RN SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT ESSA;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: Z99708; E35320; -.
DR PFAM: PF01344; Kelch; 2.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 326 AA; 36868 MW; E99CC407 CRC32;

Query Match 73.0%; Score 46; DB 10; Length 326;
Best Local Similarity 50.0%; Pred. No. 2.09e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 159 EEDOSFSF 166
|:::|::|
QY 412 EQHSFAFN 419

RESULT 11
ID O83634 PRELIMINARY; PRT: 391 AA.
AC O83634;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE EXONUCLEASE, PUTATIVE.
GN TP0626.
OS TREPONEMA PALLIDUM.
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; TREPONEMA.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 98332770.
RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA McDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RT "Complete Genome Sequence of Treponema pallidum, the Syphilis
RT Spirochete.";
RL SCIENCE 281:375-388(1998).
[2]
RN SEQUENCE FROM N.A.
RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA McDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE001237; G332921; -.
KW EXONUCLEASE.
SQ SEQUENCE 391 AA; 44297 MW; B590EB11 CRC32;

Query Match 73.0%; Score 46; DB 2; Length 391;

```

```

Best Local Similarity 62.5%; Pred. No. 2.09e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 328 QENAFADF 335
|:::|::|
QY 413 EQHSFAFN 420

RESULT 12
ID Q21093 PRELIMINARY; PRT: 529 AA.
AC Q21093;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE K01D12.2 PROTEIN.
GN K01D12.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
[1]
RN SEQUENCE FROM N.A.
RA DOBSON R.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL: Z75543; E1347664; -.
SQ SEQUENCE 529 AA; 60193 MW; E6184FB6 CRC32;

Query Match 73.0%; Score 46; DB 5; Length 529;
Best Local Similarity 55.6%; Pred. No. 2.09e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 487 EESHNEGFN 495
|:::|::|
QY 412 EQHSFAFN 420

RESULT 13
ID O87109 PRELIMINARY; PRT: 1002 AA.
AC O87109;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SURFACE LAYER PROTEIN.
GN SLAA.
OS SERRATIA MARCESCENS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SERRATIA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-SR41;
RX MEDLINE; 98194707.
RA KAWAI E., AKATSUKA H., IDEI A., SHIBATANI T., OMORI K.;
RT "Serratia marcescens S-layer protein is secreted extracellularly via
RT an ATP-binding cassette exporter, the Lip system.";
RL MOL. MICROBIOL. 27:941-952(1998).
DR EMBL: AB007124; D1034491; -.
SQ SEQUENCE 1002 AA; 101184 MW; D279246C CRC32;

```


Query Match 73.0%; Score 46; DB 2; Length 1002;
Best Local Similarity 50.0%; Pred. No. 2.09e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 383 QOHAFFD 390
Qy 413 QEHSPAFN 420

RESULT 14
ID Q54455 PRELIMINARY; PRT: 1004 AA..
AC Q54455;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SURFACE LAYER PROTEIN.
GN SLAA.
OS SERRATIA MARCESCENS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SERRATIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SR41;
RX MEDLINE: 98194707.
RA KAWAI E., AKATSUKA H., IDEI A., SHIBATANI T., OMORI K.;
RT "Serratia marcescens S-layer protein is secreted extracellularly via
an ATP-binding cassette exporter, the Lip system.";
RL MOL. MICROBIOL. 27:941-952(1998).
RN [2]
RP SEQUENCE OF 781-1004 FROM N.A.
RC STRAIN-SR41;
RX MEDLINE: 96062219.
RA AKATSUKA H., KAWAI E., OMORI K., SHIBATANI T.;
RT "The three genes lipB, lipC, and lipD involved in the extracellular
secretion of the Serratia marcescens lipase which lacks an N-terminal
signal peptide.";
RL J. BACTERIOL. 177:6381-6389(1995).
DR EMBL: AB007125; D1026234; -.
DR EMBL: D49826; D1009242; -.
SQ SEQUENCE 1004 AA; 102405 MW; 45EBA7DF CRC32;

Query Match 73.0%; Score 46; DB 2; Length 1004;
Best Local Similarity 50.0%; Pred. No. 2.09e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 385 QOHAFFD 392
Qy 413 QEHSPAFN 420

RESULT 15
ID O84567 PRELIMINARY; PRT: 94 AA.
AC O84567;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE YOPS/FLIQ TRANSLOCATION PROTEIN.
GN YSCS.
OS CHLAMYDIA TRACHOMATIS.
OC BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
Chlamydia trachomatis.";
RL SCIENCE 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,

RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE001327; G3329004; -.
SQ SEQUENCE 94 AA; 10573 MW; 24860BCA CRC32;
Query Match 71.4%; Score 45; DB 2; Length 94;
Best Local Similarity 71.4%; Pred. No. 3.35e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 51 QEQTFAP 57
Qy 413 QEHSPAF 419

Search completed: Thu Jul 8 18:53:35 1999
Job time : 15 secs.

THIS PAGE BLANK (USPTO)

WQREH (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:56:43 1999; MasPar time 9.15 Seconds
Tabular output not generated. 18.595 Million cell updates/sec

Title: >US-09-041-236-2
Description: (440-447) from US09041236.pap (24 of 45)
Perfect Score: 62
Sequence: 1 AERRKLYV 8

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 3:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 17.219; Variance 49.440; scale 0.348

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	52	83.9	1726	20 W00385	Truncated plasmodium	5.08e+01
2	52	83.9	2913	20 W00384	Plasmodium falciparum	5.08e+01
3	49	79.0	683	27 W36794	Novel human protein,	1.09e+02
4	49	79.0	785	36 W2195	HSV-2 strain SB5 Cont	1.09e+02
5	49	79.0	2182	23 W22476	Plasmodium var-1.	1.09e+02
6	49	79.0	2703	13 W0236	P. falciparum Proj3.	1.09e+02
7	49	79.0	2710	23 W22482	Plasmodium Proj3.	1.09e+02
8	49	79.0	3060	23 W22475	Plasmodium var-7.	1.09e+02
9	48	77.4	890	26 R0926	E. histolytica pyruva	1.41e+02
10	48	77.4	885	14 R0710	Recombinant cold-resi	1.41e+02
11	48	77.4	1085	18 R95607	RRP3 telomerase-assoc	1.41e+02
12	48	77.4	1106	1 P90646	Human platelet-derive	1.41e+02
13	48	77.4	1106	1 P90127	Platelet-derived grow	1.41e+02
14	48	77.4	1106	18 R95690	Platelet-derived grow	1.41e+02
15	48	77.4	1106	5 R22205	Type B human platelet	1.41e+02
16	48	77.4	1456	7 R34129	PVX replicase.	1.41e+02

17	48	77.4	2985	32 W56643	DNA-dependent protein	1.41e+02
18	48	77.4	3960	32 W56642	Protein kinase cataly	1.41e+02
19	47	75.8	601	35 W77049	Mouse striated muscle	1.81e+02
20	47	75.8	661	35 W77048	Human striated muscle	1.81e+02
21	47	75.8	817	29 W48101	Human discs large 1 g	1.81e+02
22	47	75.8	849	29 W48102	Human discs large 1 g	1.81e+02
23	47	75.8	906	27 W36795	Novel human protein,	1.81e+02
24	46	74.2	164	33 W60970	Streptococcus pneumon	2.32e+02
25	46	74.2	241	33 W55085	Streptococcus pneumon	2.32e+02
26	46	74.2	311	19 W00994	Interleukin-1 beta co	2.32e+02
27	46	74.2	323	19 W00994	ICE mutant E324X.	2.32e+02
28	46	74.2	339	19 W00989	Human interleukin-1 b	2.32e+02
29	46	74.2	339	19 R45269	ICE mutant W340X.	2.32e+02
30	46	74.2	394	7 R33779	ICE.	2.32e+02
31	46	74.2	404	9 R45270	ICE mutant A361V.	2.32e+02
32	46	74.2	404	7 R34849	Interleukin-1-beta pr	2.32e+02
33	46	74.2	404	9 R45267	ICE mutant G287S.	2.32e+02
34	46	74.2	404	19 W00985	Human interleukin-1 b	2.32e+02
35	46	74.2	404	9 R45271	ICE mutant E390K.	2.32e+02
36	46	74.2	404	9 R45272	ICE mutant T393F.	2.32e+02
37	46	74.2	404	3 R15745	Interleukin-1 beta pr	2.32e+02
38	46	74.2	404	14 R86246	Interleukin-1-beta-co	2.32e+02
39	46	74.2	404	19 W00992	Human interleukin-1 b	2.32e+02
40	46	74.2	404	19 R98755	Human interleukin-1 b	2.32e+02
41	46	74.2	404	9 R45266	ICE mutant G85R.	2.32e+02
42	46	74.2	404	9 R45263	ICE.	2.32e+02
43	46	74.2	404	19 W00987	Human interleukin-1 b	2.32e+02
44	46	74.2	404	19 W00990	Human interleukin-1 b	2.32e+02
45	46	74.2	1513	8 R43253	p190 protein.	2.32e+02

ALIGNMENTS

RESULT 1
ID W00385 standard; Protein: 1726 AA.
AC W00385;
DT 21-FEB-1997 (first entry)
DE Truncated Plasmodium falciparum erythrocyte membrane protein.
KW Plasmodium falciparum; erythrocyte membrane protein; malaria;
KW detection; identification; treatment; prevention; parasite.
OS Plasmodium falciparum MC type.
FH Key Location/Qualifiers
FT domain 62..394 /label= Duffy binding ligand domain 1
FT region 607..648 /note= "Cysteine rich motif"
FT domain 839..1282 /label= Duffy binding ligand domain 2
FT region 1488..1523 /note= "Cysteine rich motif"
PN W09633736-A1.
PD 31-OCT-1996.
PF 26-APR-1996; U05798.
PR 27-APR-1995; US-430908.
PA (AFFY-) AFFYMAX TECHNOLOGIES NV.
PI Baruch DI, Howard RJ, Pasloske BL;
DR WPI; 96-497376/49.
DR N-PSDB; T41853.
DR New Plasmodium falciparum erythrocyte membrane proteins - used to
PT develop products for the diagnosis, treatment or prevention of
PT malaria parasite infections
PS Claim 1; Figure 12; 149pp; English.
CC A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
CC membrane protein 1 (PfEMP1), or active fragments or analogues of that
CC protein can be used in the treatment or prevention of symptoms of a
CC malaria parasite infection. The polypeptides can inhibit, block or
CC reverse the sequestration of erythrocytes in patients suffering from
CC malaria. Nucleic acids derived from the PfEMP1 gene can be used as
CC probes and primers to identify a Plasmodium falciparum parasite, the
CC primers used to generate characteristic amplification patterns from
CC different P. falciparum strains. Antibodies specifically
CC immunoreactive with the PfEMP1 polypeptide or its fragments may be
CC used in diagnosis of malaria infection. This is a truncated PfEMP1

CC protein of the MC type of Plasmodium falciparum. The full length
CC PfEMP1 protein is given in W00384.
SQ Sequence 1726 AA;

Query Match 83.9%; Score 52; DB 20; Length 1726;
Best Local Similarity 100.0%; Pred. NO. 5.08e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 895 rrklyv 900
|||||
Qy 442 RRKLYV 447

RESULT 2

ID W00384 standard; Protein; 2913 AA.
AC W00384;
DT 21-FEB-1997 (first entry)
DE Plasmodium falciparum erythrocyte membrane protein.
KW Plasmodium falciparum; erythrocyte membrane protein; malaria;
KW detection; identification; treatment; prevention; parasite.
OS Plasmodium falciparum MC type.
FH Key Location/Qualifiers
FT domain 62..394
FT region /label= Duffy binding ligand domain 1
FT region 607..648
FT domain /note= "Cysteine rich motif"
FT domain 839..1272
FT domain /label= Duffy binding ligand domain 2
FT region 1482..1527
FT region /note= "Cysteine rich motif"
FT domain 1706..2005
FT domain /label= Duffy binding ligand 3
FT domain 2102..2349
FT domain /label= Duffy binding ligand 4
FT region 2354..2398
FT region /note= "Cysteine rich motif"
FT domain 2450..2475
FT domain /note= "Putative transmembrane domain"

PN W09633736-A1.

PD 31-OCT-1996.

PF 26-APR-1996; U05798.

PR 27-APR-1995; US-430908.

PA (AFFY-) AFFYMAX TECHNOLOGIES NV.

PI Baruch DI, Howard RJ, Pasloske BL;

DR WPI: 96-497376/49.

DR N-PSDB: T41852.

PT New Plasmodium falciparum erythrocyte membrane proteins - used to
PT develop products for the diagnosis, treatment or prevention of
PT malaria parasite infections

PS Claim 1; Figure 12; 149pp; English.

CC A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
CC membrane protein 1 (PfEMP1) or active fragments or analogues of that
CC protein can be used in the treatment or prevention of symptoms of a
CC malaria parasite infection. The polypeptides can inhibit, block or
CC reverse the sequestration of erythrocytes in patients suffering from
CC malaria. Nucleic acids derived from the PfEMP1 gene can be used as
CC probes and primers to identify a Plasmodium falciparum parasite, the
CC primers used to generate characteristic amplification patterns from
CC different P. falciparum strains. Antibodies specifically
CC immunoreactive with the PfEMP1 polypeptide or its fragments may be
CC used in diagnosis of malaria infection. This is the PfEMP1 protein
CC of the MC type of Plasmodium falciparum. An alternative, truncated
CC PfEMP1 protein is given in W00385.

SQ Sequence 2913 AA;

Query Match 83.9%; Score 52; DB 20; Length 2913;
Best Local Similarity 100.0%; Pred. NO. 5.08e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 892 rrklyv 897
|||||
Qy 442 RRKLYV 447

RESULT 3

ID W36794 standard; Peptide; 683 AA.

AC W36794;

DT 23-APR-1998 (first entry)

DE Novel human protein, designated WWP1, which contains WW domains.

KW Peptide recognition unit; YAP WW domain binding protein; WBP-1; WBP-2;

KW WW domain; cell signalling; growth regulation; cytoskeleton organisation;

OS targeted drug screening; modulator; WW domain interaction; WWP1.

FH Homo sapiens.

FT Key Location/Qualifiers

FT Domain 162..197

FT Domain /note= "WW domain 1"

FT Domain 192..229

FT Domain /note= "WW domain 2"

FT Domain 267..305

FT Domain /note= "WW domain 3"

FT Domain 307..344

FT Domain /note= "WW domain 4"

PN W09737223-A1.

PD 09-OCT-1997.

PF 03-APR-1997; U05547.

PR 03-APR-1996; US-630916.

PA (CYTO-) CYTOGEN CORP.

PI (UYNC-) UNIV NORTH CAROLINA.

PI Fowlkes DM, Kay BK, Pirozzi G;

DR WPI: 97-503234/46.

DR N-PSDB: T95697.

PT Identifying cell signalling and growth regulatory polypeptides by

PT reaction with multivalent recognition complex - polypeptides are

PT useful in targetted drug selection

PS Claim 48; Fig 17; 220pp; English.

CC The present sequence represents a novel protein WWP1. The WWP1 gene was

CC identified and isolated from human bone marrow and brain cDNA libraries,

CC using peptides W38103-05. These peptide recognition units are based on

CC the sequences of WW domain binding domains of the YAP WW domain binding

CC proteins WBP-1 and WBP-2. The WW domain is a small functional domain

CC found in a large number of proteins from a variety of species including

CC humans, nematodes and yeast. Its name is derived from the observation

CC that two tryptophan residues, one in the amino terminal portion of the

CC WW domain and one in the carboxyl terminal portion, are conserved. Most

CC proteins containing WW domains have a function involving cell signalling

CC and growth regulation or the organisation of the cytoskeleton.

CC Polypeptides containing a WW domain are identified by treating a

CC multivalent recognition unit complex that has selective binding affinity

CC for a WW domain, with many polypeptides and identifying those with a

CC selective affinity for the complex. Proteins containing WW domains are

CC used for targeted drug screening, i.e. to identify potential modulators

CC of specific WW domain interactions. The valency of the recognition unit

CC is important in determining specificity of interaction with WW domains.

CC In multivalent form specificity is relaxed, but not lost, so proteins

CC containing WW domains similar, but not identical, to the sequence of the

CC peptides' target WW can be detected, including new polypeptides.

SQ Sequence 683 AA;

Query Match 79.0%; Score 49; DB 27; Length 683;

Best Local Similarity 83.3%; Pred. No. 1.09e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 404 rrrlyv 409

|||||

Qy 442 RRKLYV 447

RESULT 4

ID W72195 standard; Protein; 785 AA.

AC W72195;

DT 13-JAN-1999 (first entry)

DE HSV-2 strain SB5 Contig ID 15 ORF#30 protein.

KW HSV-2 strain SB5; immunological response induction; therapy;

KW antiviral identification; viral protein inhibitor.

OS Herpes simplex virus type 2.

PN W09820016-A1.

PD 14-MAY-1998.
 PF 31-OCT-1997; U20016.
 PR 09-JUN-1997; US-049018.
 PR 04-NOV-1996; US-030279.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Chan JY, Dabrowski-Amarel CE, Delvecchio AM, Dillon SB,
 PI Esser KM, Leary JJ;
 PI WPI: 98-286847/25.
 DR N-PSDB; V62176.
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
 PT and treatment of infection or inducing immunological response in
 PT mammal
 PS Claim 10; Page 122; 748pp; English.
 CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
 CC sequence of the invention. This sequence was isolated from a HSV-2 strain
 CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.
 CC The proteins can be used for the treatment or prevention of disease, to
 CC induce an immunological response in a mammal or to identify inhibitors,
 CC activators or novel antivirals. Antagonists of the proteins can be used
 CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
 CC it can also be used to induce an immunological response in a mammal.
 SQ Sequence 785 AA;
 Query Match 79.0%; Score 49; DB 36; Length 785;
 Best Local Similarity 71.4%; Pred. No. 1.09e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 497 perrily 503
 QY 440 AERRKLY 446
 RESULT 5
 ID W22476; standard; Protein; 2182 AA.
 AC W22476;
 DT 12-SEP-1997 (first entry)
 DE Plasmodium var-1.
 KW DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium.
 OS Plasmodium vivax.
 OS Plasmodium falciparum.
 PN WO9640766-A2.
 PD 19-DEC-1996.
 PF 07-JUN-1996; U09508.
 PR 07-JUN-1995; US-487826.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR N-PSDB; T72883.
 PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and sialic acid
 PT binding proteins
 PS Claim 8; Page 70-74; 96pp; English.
 CC This sequence represents var-1 of Plasmodium. Var-1 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and sialic acid binding protein
 CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
 CC family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and SABP
 CC mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the invention.
 CC The compositions are for the treatment and prevention of malaria, and
 CC comprise either a nucleotide sequence or encoded polypeptide of the
 CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
 CC genes having homology with conserved regions of DABP and SABP. The
 CC compositions are used for the treatment and prevention of malaria. They
 CC are also used in the preparation of vaccines for inducing a protective

CC immune response in a mammal to Plasmodium merozoites (especially
 CC Plasmodium falciparum or Plasmodium vivax).
 SQ Sequence 2182 AA;
 Query Match 79.0%; Score 49; DB 23; Length 2182;
 Best Local Similarity 83.3%; Pred. No. 1.09e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 909 rrrlyv 914
 QY 442 RRKLYV 447
 RESULT 6
 ID R70236; standard; Protein; 2703 AA.
 AC R70236;
 DT 22-SEP-1995 (first entry)
 DE P. falciparum Proj3.
 KW Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;
 KW vaccine.
 OS Plasmodium falciparum.
 PN WO9507353-A.
 PD 16-MAR-1995.
 PF 07-SEP-1994; U10230.
 PR 10-SEP-1993; US-119677.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI: 95-123427/16.
 DR N-PSDB; Q83529.
 PT New erythrocyte binding domain polypeptide(s) - isolated from
 PT Plasmodium binding proteins, used in diagnosis, treatment and
 PT prevention of malaria
 PS Disclosure; Page 61-65; 81pp; English.
 CC Erythrocyte binding ligand (EBL) family genes were cloned from
 CC P. falciparum chromosome 7 subsegment libraries constructed during
 CC genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (Q83526), E3la (Q83527), EBL-e2 (Q83528) and Proj3 (Q83529),
 CC encode the proteins given in R70233-36, respectively. The binding
 CC domains of such proteins can be expressed e.g. in E. coli, yeast,
 CC mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.
 SQ Sequence 2703 AA;
 Query Match 79.0%; Score 49; DB 13; Length 2703;
 Best Local Similarity 83.3%; Pred. No. 1.09e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 1789 rrrlyv 1794
 QY 442 RRKLYV 447
 RESULT 7
 ID W22482; standard; Protein; 2710 AA.
 AC W22482;
 DT 07-OCT-1997 (first entry)
 DE Plasmodium Proj3.
 KW DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium.
 OS Plasmodium falciparum.
 PN WO9640766-A2.
 PD 19-DEC-1996.
 PF 07-JUN-1996; U09508.
 PR 07-JUN-1995; US-487826.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI: 97-052231/05.
 DR N-PSDB; T72897.
 PT New malaria vaccines - contains cysteine-rich DBL family protein

PT binding domains homologous domains of the Duffy and sialic acid
 PT binding proteins
 PS Disclosure; Page 50-56; 96pp; English.
 CC This sequence represents ProJ3 of Plasmodium. proJ3 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and sialic acid binding protein
 CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
 CC family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and SABP
 CC mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the invention.
 CC The compositions are for the treatment and prevention of malaria, and
 CC comprise either a nucleotide sequence or encoded polypeptide of the
 CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
 CC genes having homology with conserved regions of DABP and SABP. The
 CC compositions are used for the treatment and prevention of malaria. They
 CC are also used in the preparation of vaccines for inducing a protective
 CC immune response in a mammal to Plasmodium merozoites (especially
 CC Plasmodium falciparum or Plasmodium vivax).
 CC Sequence 2710 AA;

Query Match 79.0%; Score 49; DB 23; Length 2710;
 Best Local Similarity 83.3%; Pred. No. 1.09e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1796 rrrlyv 1801
 ||:||||
 Qy 442 RRKLYV 447

RESULT 8

ID W22475 standard; Protein; 3060 AA.

AC W22475;
 DT 12-SEP-1997 (first entry)
 DE Plasmodium var-7.
 KW DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium.
 OS Plasmodium vivax.
 PS Plasmodium falciparum.
 PN W09640766-A2.
 PD 19-DEC-1996.
 PF 07-JUN-1996; U09508.
 PR 07-JUN-1995; US-487826.
 PA (USHS) US DEPT HEALTH & HUMAN SERVICES.
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Welles TE;
 DR WPI: 97-052231/05.
 DR N-PSDB; T72882.

PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and sialic acid
 PT binding proteins
 PS Claim 8; Page 61-67; 96pp; English.
 CC This sequence represents var-7 of Plasmodium. Var-7 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and sialic acid binding protein
 CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
 CC family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and SABP
 CC mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the invention.
 CC The compositions are for the treatment and prevention of malaria, and
 CC comprise either a nucleotide sequence or encoded polypeptide of the
 CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
 CC genes having homology with conserved regions of DABP and SABP. The
 CC compositions are used for the treatment and prevention of malaria. They

CC are also used in the preparation of vaccines for inducing a protective
 CC immune response in a mammal to Plasmodium merozoites (especially
 CC Plasmodium falciparum or Plasmodium vivax).
 CC Sequence 3060 AA;

Query Match 79.0%; Score 49; DB 23; Length 3060;
 Best Local Similarity 83.3%; Pred. No. 1.09e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1794 rrrlyv 1799
 ||:||||
 Qy 442 RRKLYV 447

RESULT 9

ID R09026 standard; Protein; 885 AA.

AC R09026;
 DT 26-SEP-1996 (first entry)
 DE E. histolytica pyruvate phosphate dikinase.
 KW Wild type; pyruvate phosphate dikinase; PPK; cold-resistant;
 KW F. brownii; F. bidentis; maize; cold resistance.
 OS Entamoeba histolytica.
 PN W09604369-A1.
 PD 15-FEB-1996.
 PF 30-MAY-1995; J01040.
 PR 29-JUL-1994; JP-197780.
 PR 01-DEC-1994; WO-J02022.
 PA (NIBS) JAPAN TOBACCO INC.
 PI Burnell NJ, Ohta S, Usami S;
 DR WPI; 96-129386/13.
 DR N-PSDB; T12675.
 PT Recombinant cold-resistant pyruvate phosphate dikinase - imparts
 PT improved cold tolerance to transgenic plants which express the
 PT enzyme
 PS Claim 1; Page 41-47; 70pp; Japanese.
 CC The sequences given in R09023-26 represent wild type pyruvate phosphate
 CC dikinase (PPDK) enzymes from various organisms. These sequences
 CC were used in the design of a novel protein having cold-resistant
 CC PPK activity (see also R50927). The cold-resistant PPK is based
 CC primarily on the PPK sequence of F. brownii and comprises an
 CC addition, deletion or substitution of one or more amino acids in the
 CC C-terminal portion, esp. within the region residues 832-955. The
 CC cold-resistant PPK may also be based on the F. bidentis PPK.
 CC sequence having Pro at position 869 and/or Leu at 885 and Val at 952.
 CC Plants, e.g. maize, transformed with DNA encoding the cold-resistant
 CC PPK, have improved cold resistance.
 CC Sequence 885 AA;

Query Match 77.4%; Score 48; DB 18; Length 885;
 Best Local Similarity 71.4%; Pred. No. 1.41e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 329 erkklym 335
 ||:||||
 Qy 441 ERKLYV 447

RESULT 10

ID R76710 standard; Protein; 885 AA.

AC R76710;
 DT 25-JAN-1996 (first entry)
 DE Recombinant cold-resistant pyruvate phosphate dikinase.
 KW Recombinant pyruvate phosphate dikinase; cold-resistant;
 KW transformed plants; maize; deccan grass; millet; sorghum.
 OS Flaveria sp.
 PN W09515385-A1.
 PD 08-JUN-1995.
 PF 01-DEC-1994; J02022.
 PR 03-DEC-1993; AU-052275.
 PR 29-JUL-1994; JP-197780.
 PA (NIBS) JAPAN TOBACCO INC.
 PI Burnell NJ, Ohta S, Usami S;
 DR WPI; 95-215265/28.

DR N-PSDB; Q94644.
PT Recombinant cold-resistant pyruvate phosphate di-kinase - and DNA
PS coding for it which can be used to impart cold resistance to plants.
Claim 1: Pages 41-47; 70pp; Japanese.
CC Q94644 encodes R76710 a recombinant cold-resistant pyruvate phosphate
CC di-kinase (PPDK). The recombinant PPDK can be used in the prodn. of
CC transformed cold-resistant plants, e.g. maize, deccan grass, millet
CC or sorghum.
SQ Sequence 885 AA;

Query Match 77.4%; Score 48; DB 14; Length 885;
Best Local Similarity 71.4%; Pred. No. 1.41e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 329 erkkllym 335
|::|::|
QY 441 ERRKLY 447

RESULT 11
ID R95607 standard; Protein; 1085 AA.
AC R95607;
DT 12-OCT-1996 (first entry)
DE RRP3 telomerase-associated protein.
KW Yeast; RRP3 protein; telomerase-associated protein; STR7;
KW suppressor of telomeric repression-7; telomerase; ribonucleoprotein;
KW telomere; tumour; pathogen; sperm; ovum; reporter gene;
KW drug screening; antibody; immunoassay; antitumour; antiseptic;
KW contraceptive; infertility; diagnostic; gene therapy.
OS Saccharomyces cerevisiae.
PN W09612811-A2.
PD 02-MAY-1996.
PF 20-OCT-1995; U13801.
PR 20-OCT-1994; US-326781.
PR 28-APR-1995; US-431080.
PA (ARCH-) ARCH DEV CORP.
PI Gottschling DE Singer MS;
DR WPI: 96-239169/24.
DR N-PSDB; T27052.
PT Novel telomerase associated polypeptide(s) and related nucleic acid
PT - useful for detecting e.g. tumour cells or pathogens
PS Example 10: Page 294-302; 349pp; English.
CC RRP3 telomerase-associated protein (containing fragment STR7
CC (R95606) is found in conjunction with Saccharomyces cerevisiae
CC telomerase, a ribonucleoprotein required for telomere replication.
CC Other proteins associated with telomerase are given in R95601-05
CC (other STR proteins). These proteins combine with telomerase to
CC repress telomere silencing of gene expression. Oligonucleotides
CC from the encoding sequence may be used to detect non-ciliate
CC telomerase-associated genes, e.g. in tumour, pathogen, sperm or
CC ovum cells. New telomerase-associated sequences may be detected by
CC a reporter gene expression system linked to an expression-
CC repressing telomere sequence, and binding compounds, e.g.
CC antibodies, may be detected by complex formation with telomerase
CC components. The products may be used as antitumour, antiseptic or
CC contraceptive agents, in infertility diagnosis, or in gene therapy.
SQ Sequence 1085 AA;

Query Match 77.4%; Score 48; DB 18; Length 1085;
Best Local Similarity 83.3%; Pred. No. 1.41e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 474 'qrrkly 479
|::|::|
QY 441 ERRKLY 446

RESULT 12
ID P90646 standard; protein; 1106 AA.
AC P90646;
DT 20-OCT-1989 (first entry)
DE Human platelet-derived growth factor receptor.
KW Human platelet derived growth factor receptor; agonist

and antagonist drugs; wound healing; prevents atherosclerosis;
KW cancer; genetic disorders; antibodies.
OS Homo sapiens (human)
FH Key Location/Qualifiers
FT region 45..47
FT region 89..91
FT region 103..105
FT region 215..217
FT region 230..232
FT region 292..294
FT region 307..309
FT region 354..356
FT region 371..373
FT region 468..470
FT region 479..481
FT domain 531..555
PN EP-327369-A.
PD 09-AUG-1989.
PF 02-FEB-1989; 301021.
PF 02-FEB-1988; US-151414.
PA (REGC) Univ of California.
PI Williams L T; Escobedo J E.
DR WPI: 89-229378/32.
PT New DNA encoding human platelet derived growth factor receptor
PT - useful eg for assessing agonist and antagonist drugs.
PS Claim 1; page 3; 12pp; English..
CC Human platelet derived growth factor receptor (see N90388)
CC for features). Used to make probes and antibodies, and to
CC evaluate drugs. The regions specified are potential N-glycosylation
CC sites, and the domain is a transmembrane sequence.
CC Sequence 1106 AA;
SQ Sequence 1106 AA;

Query Match 77.4%; Score 48; DB 1; Length 1106;
Best Local Similarity 57.1%; Pred. No. 1.41e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 112 erkrlly 118
|::|::|
QY 441 ERRKLY 447

RESULT 13
ID P90127 standard; protein; 1106 AA.
AC P90127;
DT 1-NOV-1989 (first entry)
DE Platelet derived growth factor receptor
DE Platelet derived growth factor receptor; human; ligand binding
KW receptor analogues; isoforms; assays; antibodies; atherosclerosis;
KW wound healing; peptide dimer; cDNA.
OS Homo sapiens
PN EP-325224-A.
PD 26-JUL-1989.
PF 18-JAN-1989; 100787.
PF 22-JAN-1988; US-146877.
PA (ZYMO) Zymogenetics Inc.
PI Siedziewski AZ, Bell LA, Kindsvogel WR;
DR WPI: 89-214434/30.
DR N-PSDB; N90355.
PT Secreted ligand-binding receptor analogues eg PDGF receptor
PT - used in assays, in purifications and as, or with,
PT therapeutic agents.
PS Claim 2; fig 1; 45pp; English.
CC Platelet derived growth factor receptor (see corresp.
CC N90355). Used in the invention to make analogues that are secreted
CC (pref. Ile-29 - Met-441, and Ile-29 - Lys-531; see specification for
CC details). These analogues are easily purified, produced in large
CC quantities recombinantly, used to produce antibodies, to screen
CC ligands as imaging agents, as (ant-)agonists, or therapeutically
CC for atherosclerosis and wound healing, and assays.
SQ Sequence 1106 AA;

Query Match 77.4%; Score 48; DB 1; Length 1106;
Best Local Similarity 57.1%; Pred. No. 1.41e+02;

Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 112 erklyi 118
|::|:
QY 441 ERRLV 447

RESULT 14

ID R99690 standard; Protein; 1106 AA.
AC R99690;
DT 11-OCT-1996 (first entry)
DE Platelet-derived growth factor receptor.
KW Platelet-derived growth factor receptor; PDGF-R; peptide dimer;
KW protein secretion; agonist; antagonist.
OS Homo sapiens.
PN EP-721983-A1.
PD 17-JUL-1996.
PF 18-JAN-1989; 100787.
PR 22-JAN-1988; US-146877.
PI (ZIMO) ZYMOGENETICS INC.
PI Bell LA, Kindsvogel WR, Sledziewski AZ;
DR WPI: 96-322833/33.
DR N-PSDB: T34552
PT Prodn. of biologically active peptide dimers, esp. platelet-derived
PT growth factor receptor analogues - useful for systematic designing
PT of novel (ant)agonists
PS Disclosure; Fig 1; 45pp; English.
CC The amino acid sequence (R99690) of human platelet-derived growth
CC factor receptor (PDGF-R) was deduced from a cDNA clone (T34552)
CC isolated from a human diploid dermal fibroblast library. The
CC cDNA can be used in novel constructs that allow the prodn.
CC of secreted biologically active PDGF-R analogues. This may
CC comprise linking a sequence coding for PDGF-R, or the ligand-
CC binding domain of the PDGF-R extracellular region, to a protein
CC secretion signal (e.g. SUC2) and promoter, and expression in
CC transformed host cells, esp. Saccharomyces cerevisiae. The
CC secreted PDGF-R analogues are used in ligand screening procedures,
CC to screen for (ant)agonists, and in diagnostic assays.
SQ Sequence 1106 AA;

Query Match 77.4%; Score 48; DB 18; Length 1106;
Best Local Similarity 57.1%; Pred. No. 1.41e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 112 erklyi 118
|::|:
QY 441 ERRLV 447

RESULT 15

ID R26205 standard; Protein; 1106 AA.
AC R26205;
DT 09-FEB-1993 (first entry)
DE Type B human platelet-derived growth factor receptor.
KW PDGF; PDGF-R; mesenchyme; tyrosine kinase; ligand binding region.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..32
FT /label= Signal_peptide
FT 24..1106
FT /label= Mature_PDGf-B
PN WO9213867-A.
PD 20-AUG-1992.
PF 28-JAN-1992; U00730.
PR 31-JAN-1991; US-650793.
PA (CORT-) COR THERAPEUTICS INC.
PI Escobedo JA, Fretto LJ, Giese NA, Tomlinson JE, Williams LT;
PI Wolf D;
DR WPI: 92-299970/36.
DR N-PSDB: Q27447.
DR Platelet derived growth factor receptor (PDGF-R) poly:peptide(s)
PT - uncluf as therapeutic and diagnostic agents e.g. for assaying
PT PDGF activity in sample

PS Disclosure; Page 75; 109pp; English.
CC The sequence given is one allele of type B human platelet-derived
CC growth factor (PDGF) receptor (PDGF-R). This receptor is typically
CC found on cells of mesenchymal origin. It acts while in the form of
CC two transmembrane glycoproteins, each of which is about 180 kD.
CC This receptor has three major regions. The first is a transmembrane
CC region, which spans the membrane once, separating the regions of the
CC receptor exterior to the cell from those interior to the cell. The
CC second region is an extracellular region which contains the domains
CC which bind the PDGF. The third region is an intracellular region
CC which possesses a tyrosine kinase activity. This tyrosine kinase
CC domain is notable in having an insert of approx. 100 amino acids,
CC as compared with most other receptor tyrosine kinase domains which
CC are contiguous or have shorter insert sequences. Fragments of this
CC sequence between 8 and 400 amino acids comprising one or more PDGF
CC ligand binding region from the extracellular domain may be used to
CC bind a PDGF ligand.
SQ Sequence 1106 AA;

Query Match 77.4%; Score 48; DB 5; Length 1106;
Best Local Similarity 57.1%; Pred. No. 1.41e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 112 erklyi 118
|::|:
QY 441 ERRLV 447

Search completed: Thu Jul 8 18:57:01 1999
Job time : 18 secs.

WQELH
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:56:07 1999; Maspar time 3.63 Seconds
Tabular output not generated.
88.348 Million cell updates/sec

Title: >US-09-041-236-2
Description: (440-447) from US09041236.pap (24 of 45)
Perfect Score: 62
Sequence: 1 AERRKLYV 8

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.843; Variance 30.827; scale 0.773

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	55	88.7	363	1	WZBE44 gene 44 protein - hum	1.38e+00
2	55	88.7	663	2	A35087 arachidonate 12-lipo	1.38e+00
3	55	88.7	663	2	S32825 arachidonate 12-lipo	1.38e+00
4	52	83.9	359	2	S70632 zinc transporter Znr-	5.47e+00
5	52	83.9	418	2	S24612 creatine kinase (EC 2	5.47e+00
6	52	83.9	456	2	S14924 hypothetical protein	8.56e+00
7	51	82.3	174	2	S68246 phosphatidylcholine t	8.56e+00
8	51	82.3	662	1	A31349 arachidonate 15-lipo	8.56e+00
9	51	82.3	663	2	QJ0018 arachidonate 15-lipo	8.56e+00
10	50	80.6	249	2	S64798 ribonucleoprotein 1	1.33e+01
11	50	80.6	417	2	A30789 creatine kinase (EC 2	1.33e+01
12	50	80.6	848	2	S44732 b0523.5 protein - Cae	1.33e+01
13	50	80.6	1100	1	DBBYD1 RAD1 protein - yeast	1.33e+01
14	49	79.0	101	2	S37929 hypothetical protein	2.06e+01
15	49	79.0	379	2	S55900 DNAJ-like protein hom	2.06e+01
16	49	79.0	418	2	S17189 creatine kinase (EC 2	2.06e+01
17	49	79.0	571	1	RNCW7H transcription initiat	2.06e+01
18	49	79.0	602	2	S58336 probable membrane pro	2.06e+01
19	49	79.0	2197	2	B71600 PfEMP1 PFBI055c - mal	2.06e+01
20	48	77.4	87	2	S03934 colicin D immunity pr	3.16e+01
21	48	77.4	87	2	S03255 colicin D immunity pr	3.16e+01
22	48	77.4	182	2	D45683 helicase homolog uvrD	3.16e+01
23	48	77.4	183	2	A41302 HMG-1 protein homolog	3.16e+01

24 48 77.4 239 2 S04966 prolactin-like protei 3.16e+01
25 48 77.4 385 2 E71014 probable dipeptide tr 3.16e+01
26 48 77.4 407 2 S27774 transforming protein 3.16e+01
27 48 77.4 499 1 B43684 RNA-directed RNA poly 3.16e+01
28 48 77.4 659 2 S20313 peroxisomal membrane 3.16e+01
29 48 77.4 780 1 WMBEH8 infected cell protein 3.16e+01
30 48 77.4 785 1 WMBEK8 infected cell protein 3.16e+01
31 48 77.4 785 1 WMBEK8 infected cell protein 3.16e+01
32 48 77.4 798 2 S62031 vacuolar protein sort 3.16e+01
33 48 77.4 885 2 S36601 pyruvate,orthophospha 3.16e+01
34 48 77.4 1085 2 S55352 IFH1 protein - yeast 3.16e+01
35 48 77.4 1106 1 PFHUGB platelet-derived grow 3.16e+01
36 48 77.4 1456 2 JQ2294 hypothetical 165.1k p 3.16e+01
37 48 77.4 1456 1 WMBGPV RNA-directed RNA poly 3.16e+01
38 48 77.4 4096 2 A57099 DNA-activated protein 3.16e+01
39 47 75.8 110 2 E71151 hypothetical protein 4.83e+01
40 47 75.8 148 2 S30280 hPCR protein - Escher 4.83e+01
41 47 75.8 317 2 B37388 probable DNA-binding 4.83e+01
42 47 75.8 317 2 A37388 probable DNA-binding 4.83e+01
43 47 75.8 804 1 A69309 probable heavy-metal- 4.83e+01
44 47 75.8 926 2 A54142 nucleoporin NUP107 - 4.83e+01
45 47 75.8 1045 2 S60571 integrin alpha v chai 4.83e+01

ALIGNMENTS

RESULT 1
ENTRY WZBE44 #type complete
TITLE gene 44 protein - human herpesvirus 3
ORGANISM #formal_name human herpesvirus 3, varicella-zoster virus
DATE 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 14-Nov-1997
ACCESSIONS I27341
REFERENCE A27345
#authors Davison, A.J.; Scott, J.E.
#journal J. Gen. Virol. (1986) 67:1759-1816
#title The complete DNA sequence of varicella-zoster virus.
#cross-references MUID:86306657
#accession I27341
#molecule_type DNA
#residues 1-363 #label DAV
#cross-references EMBL:X04370; NID:g59989; PID:g60033
GENETICS
#gene 44
CLASSIFICATION #superfamily varicella-zoster virus gene 44 protein
SUMMARY #length 363 #molecular-weight 40244 #checksum 258

Query Match 88.7%; Score 55; DB 1; Length 363;
Best Local Similarity 85.7%; Pred. No. 1.38e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 187 ERRRLYV 193
|||:|
Qy 441 ERRKLYV 447

RESULT 2
ENTRY A35087 #type complete
TITLE arachidonate 12-lipoxygenase (EC 1.13.11.31) - pig
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 03-Aug-1990 #sequence_revision 31-Jan-1997 #text_change 31-Jan-1997
ACCESSIONS A35087
REFERENCE A35087
#authors Yoshimoto, T.; Suzuki, H.; Yamamoto, S.; Takai, T.; Yokoyama, C.; Tanabe, T.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:2142-2146
#title Cloning and sequence analysis of the cDNA for arachidonate 12-lipoxygenase of porcine leukocytes.
#cross-references MUID:90192763
#accession A35087
#status preliminary
#molecule_type mRNA

```
##residues 1-663 ##label YOS
##cross-references GB:M31417
CLASSIFICATION #superfamily arachidonate 5-lipoxygenase
KEYWORDS oxidoreductase
SUMMARY #length 663 #molecular-weight 74974 #checksum 4092

Query Match 88.7%; Score 55; DB 2; Length 663;
Best Local Similarity 100.0%; Pred. No. 1.38e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 134 AERRKLY 140
|||||
Qy 440 AERRKLY 446

RESULT 3
ENTRY S32825 #type complete
TITLE arachidonate 12-lipoxygenase (EC 1.13.11.31), tracheal -
ORGANISM bovine
#formal_name Bos primigenius taurus #common_name cattle
DATE 13-Jan-1995 #sequence_revision 29-Aug-1997 #text_change
08-Sep-1997
ACCESSIONS S32825; S77975; A56770
REFERENCE A56770
#authors de Marzo, N.; Sloane, D.L.; Dicharry, S.; Highland, E.;
Sigal, E.
#journal Am. J. Physiol. (1992) 262:L198-L207
#title Cloning and expression of an airway epithelial
12-lipoxygenase.
#cross-references MUID:92170942
#contents tracheal epithelium
#accession S32825
#molecule_type mRNA
##residues 1-663 ##label ACC
##cross-references EMBL:S96247; NID:9246172; PID:9246173
#note in the authors' translation residues 441-460 do not
match the nucleotide sequence
#note sequence extracted from NCBI backbone (NCBI:96247)
REFERENCE S77975
#authors de Marzo, N.; Sloane, D.L.; Dicharry, S.; Highland, E.;
Sigal, E.
#submission submitted to the EMBL Data Library, December 1993
#accession S77975
#molecule_type mRNA
##residues 1-440, 'GLLVKSSFY', 441-450, 461-663 ##label DEM
##cross-references EMBL:S96247
CLASSIFICATION #superfamily arachidonate 5-lipoxygenase
KEYWORDS leukotriene biosynthesis; oxidoreductase
SUMMARY #length 663 #molecular-weight 75041 #checksum 3031

Query Match 88.7%; Score 55; DB 2; Length 663;
Best Local Similarity 100.0%; Pred. No. 1.38e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 134 AERRKLY 140
|||||
Qy 440 AERRKLY 446

RESULT 4
ENTRY S70632 #type complete
TITLE zinc transporter Znt-2 - rat
ORGANISM Rattus norvegicus #common_name Norway rat
DATE 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change
05-Dec-1997
ACCESSIONS S70632
REFERENCE S70632
#authors Palmiter, R.D.; Cole, T.B.; Findley, S.D.
#journal EMBO J. (1996) 15:1784-1791
#title Znt-2, a mammalian protein that confers resistance to zinc by
facilitating vesicular sequestration.
#cross-references MUID:96203098
#accession S70632

##status preliminary
##molecule_type mRNA
##residues 1-359 ##label PAL
##cross-references EMBL:U50927; NID:g1256377; PID:g1256378
GENETICS
#gene Znt-2
#start_codon CTG
CLASSIFICATION #superfamily zinc transporter Znt-2
SUMMARY #length 359 #molecular-weight 39276 #checksum 2497

Query Match 83.9%; Score 52; DB 2; Length 359;
Best Local Similarity 100.0%; Pred. No. 5.47e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 54 RRKLYV 59
|||||
Qy 442 RRKLYV 447

RESULT 5
ENTRY S24612 #type complete
TITLE creatine kinase (EC 2.7.3.2), mitochondrial - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
24-Jul-1997
ACCESSIONS I48308; I48309; S24612
REFERENCE I48308
#authors Steeghs, K.; Peters, W.; Bruckwilder, M.; Croes, H.; Van
Alewijk, D.; Wieringa, B.
#journal DNA Cell Biol. (1995) 14:539-553
#title Mouse ubiquitous mitochondrial creatine kinase: gene
organization and consequences from inactivation in mouse
embryonic stem cells.
#cross-references MUID:95322032
#accession I48308
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-418 ##label RES
##cross-references EMBL:Z13969; NID:9634057; PID:9663017
#accession I48309
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-418 ##label RE2
##cross-references EMBL:Z13968; NID:9634058; PID:950562
GENETICS
#genome nuclear
#introns 51/2; 117/3; 149/3; 223/3; 252/2; 293/3; 338/3; 380/3
CLASSIFICATION #superfamily creatine kinase; creatine kinase repeat homology
mitochondrion; phosphotransferase
FEATURES
54-413
317
SUMMARY #domain creatine kinase repeat homology #label CKR\
#active_site Cys #status predicted
#length 418 #molecular-weight 47003 #checksum 5779

Query Match 83.9%; Score 52; DB 2; Length 418;
Best Local Similarity 85.7%; Pred. No. 5.47e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 42 AERRRLY 48
|||||
Qy 440 AERRRLY 446

RESULT 6
ENTRY S14924 #type complete
TITLE hypothetical protein 456 (tufa 3' region) - euglenid (Astasia
longa) plastid
ORGANISM #formal_name plastid Astasia longa
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
08-Sep-1997
ACCESSIONS S14924; S38613; S08116
REFERENCE S14920
#authors Siemeister, G.; Buchholz, C.; Hachtel, W.
```

```
#journal Mol. Gen. Genet. (1990) 220:425-432
#title Genes for the plasmid elongation factor Tu and ribosomal
      protein S7 and six tRNA genes on the 73 kb DNA from Astasia
      longa that resembles the chloroplast DNA of Euglena.
#cross-references EMBL:90251252
#accession SI4924
##molecule_type DNA
##residues 1-456 ##label SIE
##cross-references EMBL:X14385; NID:g11197; PID:g11201
REFERENCE S38590
#authors Gockel, G.; Baier, S.; Hachtel, W.
#submission submitted to the EMBL Data Library, November 1993
#accession S38613
##molecule_type DNA
##residues 1-456 ##label GOC
##cross-references EMBL:X75653; NID:9414919; PID:9414924
CLASSIFICATION #superfamily Euglena psbc intron protein
KEYWORDS chloroplast; plastid
SUMMARY #length 456 #molecular-weight 54395 #checksum 9856

Query Match      83.98; Score 52; DB 2; Length 456;
Best Local Similarity 71.4%; Pred. No. 5.47e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 37 DRKKLYV 43
QY 441 ERRKLYV 447

RESULT 7
ENTRY #type fragment
TITLE phosphatidylcholine transfer protein - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 03-Dec-1996 #sequence_revision 07-Feb-1997 #text_change
      08-Sep-1997
ACCESSIONS S68246; S57661
REFERENCE S68244
#authors Geijtenbeek, T.B.H.; Smith, A.J.; Borst, P.; Wirtz, K.W.A.
#journal Blochem. J. (1996) 316:49-55
#title cDNA cloning and tissue-specific expression of the
      phosphatidylcholine transfer protein gene.
#accession S68246
##status translation not shown
##molecule_type mRNA
##residues 1-174 ##label GEI
##cross-references EMBL:Z50024; NID:9897785; PID:9897786
CLASSIFICATION #superfamily phosphatidylcholine-transfer protein
SUMMARY #length 174 #checksum 3024

Query Match      82.3%; Score 51; DB 2; Length 174;
Best Local Similarity 71.4%; Pred. No. 8.56e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 86 DRKKLYV 92
QY 441 ERRKLYV 447

RESULT 8
ENTRY #type complete
TITLE arachidonate 15-lipoxygenase (EC 1.13.11.33) - human
ALTERNATE_NAMES arachidonate omega-6 lipoxygenase
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31-Mar-1990 #sequence_revision 07-Feb-1997 #text_change
      07-Feb-1997
ACCESSIONS A31349; A28192; S19625; A61164; B61164
REFERENCE A31349
#authors Sigal, E.; Craik, C.S.; Highland, E.; Grunberger, D.;
      Costello, L.L.; Dixon, R.A.F.; Nagel, J.A.
#journal Biochem. Biophys. Res. Commun. (1988) 157:457-464
#title Molecular cloning and primary structure of human
      15-lipoxygenase.
#cross-references MUID:89076270
```

```
#accession A31349
##molecule_type mRNA
##residues 1-662 ##label SIG
##cross-references GB:M23892; NID:g187190; PID:g307135
##experimental_source reticulocyte
REFERENCE A28192
#authors Sigal, E.; Grunberger, D.; Craik, C.S.; Caughey, G.H.; Nagel,
      J.A.
#journal J. Biol. Chem. (1988) 263:5328-5332
#title Arachidonate 15-lipoxygenase (omega-6 lipoxygenase) from
      human leukocytes. Purification and structural homology to
      other mammalian lipoxygenases.
#cross-references MUID:88186828
#accession A28192
##molecule_type protein
##residues 3-16 ##label SI2
##experimental_source leukocyte
REFERENCE S19577
#authors Izumi, T.; Radmark, O.; Joernvall, H.; Samuelsson, B.
#journal Eur. J. Biochem. (1991) 202:1231-1238
#title Purification of two forms of arachidonate 15-lipoxygenase
      from human leukocytes.
#cross-references MUID:92111501
#accession S19625
##molecule_type protein
##residues 'X', 3-4, 'X', 6, 'X', 8-22; 38-45; 157-162, 'XX', 165-168;
      627-631 ##label IZU
#accession S19577
##molecule_type protein
##residues 'X', 3-25, 27-31 ##label IZ1
##note there appear to be distinct chromatographic forms, at
      least one each from reticulocyte and eosinophil, and
      two from leukocyte, arising from a post-translational
      modification of a common precursor
REFERENCE A61164
#authors Izumi, T.; Radmark, O.; Samuelsson, B.
#journal Adv. Prostaglandin Thromboxane Leukotriene Res. (1990)
      21:101-104
#title Purification of 15-lipoxygenase from human leukocytes,
      evidence for the presence of isozymes.
#accession A61164
##molecule_type protein
##residues 'X', 3-4, 'X', 6, 'X', 8-12, 'X', 14-19, 'T', 21-22 ##label IZ2
##experimental_source leukocyte
#accession B61164
##molecule_type protein
##residues 'X', 3-25, 27-31 ##label IZ3
##experimental_source leukocyte
GENETICS
#gene GDB:ALOX15
##cross-references GDB:132454
#map_position 17pter-17qter
FUNCTION
#description catalyzes the oxidation of arachidonic acid to (5Z,8Z,11Z,
      13E)-(15S)-15-hydroperoxyicoso-5,8,11,13-tetraenoic acid
      by dioxygen
#pathway leukotriene biosynthesis
CLASSIFICATION #superfamily arachidonate 5-lipoxygenase
KEYWORDS fatty acid oxidation; iron; leukotriene biosynthesis;
      metalloprotein; oxidoreductase
FEATURE
      360,365,540,544,
      662
#binding_site iron (His, His, His, His, Ile) #status
      predicted
SUMMARY #length 662 #molecular-weight 74804 #checksum 2368

Query Match      82.3%; Score 51; DB 1; Length 662;
Best Local Similarity 100.0%; Pred. No. 8.56e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 134 ERRKLYV 139
QY 441 ERRKLYV 446
```



```

REFERENCE      A31431
#authors       Haas, R.C.; Korenfeld, C.; Zhang, Z.; Perryman, B.; Roman,
#journal       D.; Struss, A.W.
#title         J. Biol. Chem. (1989) 264:2890-2897
#cross-references MUID:89123390
#accession     A31431
#status        Preliminary
##molecule_type DNA
##residues     1-417 #label HAA
##cross-references GB:J04469; NID:g180589; PID:g180590
GENETICS
#gene          GDB:CKMT1; CKMT; UMTCK
#map_position  15q15-15q15
CLASSIFICATION #superfamily creatine kinase; creatine kinase repeat homology
KEYWORDS       mitochondrial; phosphotransferase
FEATURE
53-412         #domain creatine kinase repeat homology #label CKR\
316            #active_site Cys #status predicted
SUMMARY        length 417 #molecular-weight 47036 #checksum 6770
Query Match    80.6%; Score 50; DB 2; Length 417;
Best Local Similarity 71.4%; Pred. No. 1.33e+01;
Matches        5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 41 SERRLY 47
QY 440 AERRKLY 446

RESULT 12
ENTRY   S44732      #type complete
TITLE   b0523.5 protein - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE     20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
23-Aug-1997
ACCESSIONS S44732
REFERENCE   S44732
#authors   Wilson, R.
#submission submitted to the EMBL Data Library, December 1992
#description Sequence of the C. elegans cosmid B0523.
#accession S44732
##status   Preliminary
##molecule_type DNA
##residues 1-848 #label WIL
##cross-references EMBL:L07143
GENETICS
#introns   224/3; 545/3; 606/2; 723/3; 823/1
CLASSIFICATION #superfamily gelsolin repeat homology
FEATURE       #domain gelsolin repeat homology #label GEL1
56-387        #length 848 #molecular-weight 98495 #checksum 5926
SUMMARY
Query Match    80.6%; Score 50; DB 2; Length 848;
Best Local Similarity 85.7%; Pred. No. 1.33e+01;
Matches        6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 286 AERRKLY 292
QY 440 AERRKLY 446

RESULT 13
ENTRY   DBVD1       #type complete
TITLE   RAD1 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein LPB9w; protein YPL022w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
12-Dec-1997
ACCESSIONS A26129; S05878; S63460
REFERENCE   A26129

```

```

#authors       Reynolds, P.; Prakash, L.; Prakash, S.
#journal       Mol. Cell. Biol. (1987) 7:1012-1020
#title         Nucleotide sequence and functional analysis of the RAD1 gene
#cross-references MUID:87172766
#accession     A26129
##molecule_type DNA
##residues     1-1100 #label REV
##cross-references EMBL:M15435; NID:g172324; PID:g172325
REFERENCE      S05878
#authors       Yang, E.; Friedberg, E.C.
#journal       Mol. Cell. Biol. (1984) 4:2161-2169
#title         Molecular cloning and nucleotide sequence analysis of the
Saccharomyces cerevisiae RAD1 gene.
#cross-references MUID:85061207
#accession     S05878
##molecule_type DNA
##residues     1-222, 'N', 224-882, 'Y', 884-885, 'I', 887-911, 'K', 913-922,
'NYSSVIKQVIG', 935-947, 'SHEKDYMVLTANCKYNPRVEIRT',
'#label VAN
##cross-references EMBL:K02070; NID:g172317; PID:g172318
REFERENCE      S63452
#authors       Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.;
Hall, J.; Storms, R.K.; Vo, D.H.; Winnett, E.
#submission submitted to the EMBL Data Library, September 1995
#accession     S63460
##molecule_type DNA
##residues     1-1100 #label WAN
##cross-references EMBL:U36624; NID:g1276642; PID:g1039455; MIPS:YPL022w
GENETICS
#gene          SGD:RAD1
##cross-references SGD:S0005943; MIPS:YPL022w
#map_position  16L
CLASSIFICATION #superfamily RAD1 protein
KEYWORDS       DNA binding; DNA repair
SUMMARY        length 1100 #molecular-weight 126369 #checksum 1360
Query Match    80.6%; Score 50; DB 1; Length 1100;
Best Local Similarity 83.3%; Pred. No. 1.33e+01;
Matches        5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 203 RRKLYI 208
QY 442 RRKLYV 447

RESULT 14
ENTRY   S37929      #type complete
TITLE   hypothetical protein YKL102c - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES hypothetical protein YKL454
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     03-May-1994 #sequence_revision 03-May-1994 #text_change
17-Mar-1999
ACCESSIONS S37929; S39102
REFERENCE   S37920
#authors       Charet, G.; Fukuhara, H.; Bolotin-Fukuhara, M.;
Daignan-Fornier, B.; Pallier, C.; Puzos, V.; Valens, M.
#submission submitted to the Protein Sequence Database, March 1994
#accession     S37929
##molecule_type DNA
##residues     1-101 #label CHR
##cross-references EMBL:Z28102; NID:g486170; PID:g486171; MIPS:YKL102c
#experimental_source strain S288C
REFERENCE      S39094
#authors       Charet, G.; Pallier, C.; Valens, M.; Daignan-Fornier, B.;
Fukuhara, H.; Bolotin-Fukuhara, M.; Sor, F.
#journal       Yeast (1993) 9:1259-1265
#title         The DNA sequence analysis of the HAP4-LAP4 region on
chromosome XI of Saccharomyces cerevisiae suggests the
presence of a second aspartate aminotransferase gene in
yeast.
#cross-references MUID:94152173

```

```
#accession S39102
#status translation not shown
#molecule_type DNA
#residues 1-101 #label CHE
#cross-references EMBL:X71133; NID:9431205; PID:9431214
#experimental_source strain S288C
GENETICS
#map_position 11L
SUMMARY #length 101 #molecular-weight 11423 #checksum 5075
Query Match 79.0%; Score 49; DB 2; Length 101;
Best Local Similarity 77.8%; Pred. No. 2.06e+01;
Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
Db 34 SERRKVLV 42
QY 440 AERRK-LV 447
:|||||
:|||||

RESULT 15
ENTRY S55900 #type complete
TITLE DnaJ-like protein homolog - fission yeast
ORGANISM (Schizosaccharomyces pombe)
DATE #formal_name Schizosaccharomyces pombe
28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
21-Aug-1998
ACCESSIONS S55900
REFERENCE S55900
#authors Park, S.K.; Chon, S.K.; Yoo, H.S.
#journal Biochim. Biophys. Acta (1995) 1262:87-90
#title A cDNA of Schizosaccharomyces pombe encoding a homologue of
DnaJ-like protein.
#cross-references MUID:95290501
#accession S55900
#status preliminary
#molecule_type mRNA
#residues 1-379 #label PAR
#cross-references EMBL:L37753; NID:g576932; PID:g953212
GENETICS
#gene psi
CLASSIFICATION #superfamily heat shock protein dnaJ; dnaJ amino-terminal
homology
FEATURE
6-68 #domain dnaJ amino-terminal homology #label DNJ
SUMMARY #length 379 #molecular-weight 40306 #checksum 6737
Query Match 79.0%; Score 49; DB 2; Length 379;
Best Local Similarity 71.4%; Pred. No. 2.06e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 58 PORRKL 64
QY 440 AERRKL 446
:|||||
:|||||

Search completed: Thu Jul 8 18:56:25 1999
Job time : 18 secs.
```

WQERLH (TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:54:58 1999; MasPar time 3.06 Seconds
Tabular output not generated. 73.811 Million cell updates/sec

Title: >US-09-041-236-2
Description: (440-447) from US09041236.pgp (24 of 45)
Perfect Score: 62
Sequence: 1 AERRKLYV 8

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: swiss-prot37
1:swissprot

Statistics: Mean 24.519; Variance 27.096; scale 0.905

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description	Pred. No.
1	55	88.7	UL16_VZVD	GENE 44 PROTEIN.	3.19e-01
2	55	88.7	LOX2_PIG	ARACHIDONATE 12-LIPOXY	3.19e-01
3	55	88.7	LOX2_BOVIN	ARACHIDONATE 12-LIPOXY	3.19e-01
4	52	83.9	KCRU_MOUSE	CREATINE KINASE, UBIQU	1.53e+00
5	52	83.9	YC13_ASTLO	HYPOTHETICAL 54.4 KD P	1.53e+00
6	52	83.9	HPS_HUMAN	HERMANSKY-PUDLAK SYNDR	1.53e+00
7	51	82.3	PPCT_MOUSE	PHOSPHATIDYLCHOLINE TR	2.54e+00
8	51	82.3	LOX1_HUMAN	ARACHIDONATE 15-LIPOXY	2.54e+00
9	51	82.3	LOX1_RABIT	ARACHIDONATE 15-LIPOXY	2.54e+00
10	50	80.6	RNP1_YEAST	RIBONUCLEOPROTEIN-1.	4.19e+00
11	50	80.6	KCRU_HUMAN	CREATINE KINASE, UBIQU	4.19e+00
12	50	80.6	GBA4_USTMA	GUANINE NUCLEOTIDE-BIN	4.19e+00
13	50	80.6	RAD1_YEAST	DNA REPAIR PROTEIN RAD	4.19e+00
14	50	80.6	FLH1_CAEEL	FLIGHTLESS-1 PROTEIN H	4.19e+00
15	49	79.0	YKX2_YEAST	HYPOTHETICAL 11.4 KD P	6.85e+00
16	49	79.0	VATX_NEUCR	VACUOLAR ATP SYNTHASE	6.85e+00
17	49	79.0	PSI_SCHPO	PSI PROTEIN.	6.85e+00
18	49	79.0	KCRU_RAT	CREATINE KINASE, UBIQU	6.85e+00
19	49	79.0	STCB_EMENI	PROBABLE STERIGMATOCYS	6.85e+00
20	48	77.4	IMMD_ECOLI	COLICIN D IMMUNITY PRO	1.11e+01
21	48	77.4	UVRD_MYCCA	PUTATIVE DNA HELICASE	1.11e+01
22	48	77.4	ABF2_YEAST	ARS-BINDING FACTOR 2 P	1.11e+01
23	48	77.4	PRR4_BOVIN	PLACENTAL PROLACTIN-RE	1.11e+01

24	48	77.4	407	1	MYC_ASTVU	MYC PROTEIN (C-MYC) (F	1.11e+01
25	48	77.4	555	1	MCRA_METFE	METHYL-COENZYME M REDU	1.11e+01
26	48	77.4	767	1	RRPO_RCNMV	PUTATIVE RNA-DIRECTED	1.11e+01
27	48	77.4	780	1	PRTP_HSV1F	PROCESSING AND TRANSPO	1.11e+01
28	48	77.4	785	1	PRTP_HSV1A	PROCESSING AND TRANSPO	1.11e+01
29	48	77.4	785	1	PRTP_HSV1I	PROCESSING AND TRANSPO	1.11e+01
30	48	77.4	798	1	VP16_YEAST	VACUOLAR PROTEIN SORTI	1.11e+01
31	48	77.4	885	1	PODK_ENTHI	PYRUVATE, PHOSPHATE DK	1.11e+01
32	48	77.4	1085	1	IFH1_YEAST	IFH1 PROTEIN (RRP3 PRO	1.11e+01
33	48	77.4	1106	1	PGDR_HUMAN	BETA PLATELET-DERIVED	1.11e+01
34	48	77.4	1456	1	RRPO_PVXX3	RNA REPLICATION PROTEI	1.11e+01
35	48	77.4	1456	1	RRPO_PVXHB	RNA REPLICATION PROTEI	1.11e+01
36	47	75.8	148	1	HPCR_ECOLI	HOMOPROTEOCAEUATE DE	1.79e+01
37	47	75.8	817	1	SP02_HUMAN	PRESYNAPTIC PROTEIN SA	1.79e+01
38	47	75.8	849	1	SP02_MOUSE	PRESYNAPTIC PROTEIN SA	1.79e+01
39	47	75.8	849	1	SP02_RAT	PRESYNAPTIC PROTEIN SA	1.79e+01
40	47	75.8	926	1	N107_RAT	NUCLEAR PORE COMPLEX P	1.79e+01
41	47	75.8	1081	1	SPS2_CRAPL	SUCROSE-PHOSPHATE SYNT	1.79e+01
42	46	74.2	233	1	Y308_METJA	HYPOTHETICAL PROTEIN M	2.86e+01
43	46	74.2	708	1	YNZB_CAEEL	HYPOTHETICAL 79.0 KD P	2.86e+01
44	46	74.2	754	1	CHLD_PEA	MAGNESIUM-CHELATASE SU	2.86e+01
45	46	74.2	1137	1	MSB1_YEAST	MORPHOGENESIS-RELATED	2.86e+01

ALIGNMENTS

RESULT 1
ID UL16_VZVD STANDARD; PRT; 363 AA.
AC P09293;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE GENE 44 PROTEIN.
GN 44.
OS VARICELLA-ZOSTER VIRUS (STRAIN DUMAS) (VZV).
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPEVIRINAE; VARICELLOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86306657.
RA DAVISON A.J., SCOTT J.E.;
RT "The complete DNA sequence of varicella-zoster virus."
RL J. GEN. VIROB. 67:1759-1816(1986).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL16,
CC HSV-6 ORF1R, EBV-1 46, HCMV UL34, EBV BGLF2, HSV 33, AND VZV 44.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04370; G60033; -.
DR PIR; I27341; W2BE44.
SQ SEQUENCE 363 AA; 40244 MW; 5700C905 CRC32;

Query Match 88.7%; Score 55; DB 1; Length 363;
Best Local Similarity 85.7%; Pred. No. 3.19e-01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 187 ERRRLYV 193
|||:|
Qy 441 ERRKLYV 447

RESULT 2
ID LOX2_PIG STANDARD; PRT; 662 AA.
AC P16469;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)

```
DE ARACHIDONATE 12-LIPOXYGENASE (EC 1.13.11.31) (12-LOX).
GN ALOX12.
OS SUS SCROFA (PIG).
OC EUKARYOTA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-LEUKOCYTE;
RX MEDLINE; 90192763.
RA YOSHIMOTO T., SUZUKI H., YAMAMOTO S., TAKAI T., YOKOYAMA C.,
RT TANABE T.:
RT "Cloning and sequence analysis of the cDNA for arachidonate 12-
RT lipxygenase of porcine leukocytes.";
RL PROC. NATL. ACAD. SCI. U.S.A. 87:2142-2146(1990).
RN [2]
RN MUTAGENESIS.
RX MEDLINE; 94137772.
RA SUZUKI H., KISHIMOTO K., YOSHIMOTO T., YAMAMOTO S., KANAI F.,
RA EBINA Y., MIYATAKE A., TANABE T.:
RT "Site-directed mutagenesis studies on the iron-binding domain and the
RT determinant for the substrate oxygenation site of porcine leukocyte
RT arachidonate 12-lipoxygenase.";
RL BIOCHIM. BIOPHYS. ACTA 1210:308-316(1994).
CC -1- FUNCTION: OXYGENASE AND 14,15-LEUKOTRIENE A4 SYNTHASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ARACHIDONATE + O(2) -> (5Z,8Z,10E,14Z)-(12S)-
CC 12-HYDROPEROXYICOSA-5,8,10,14-TETRAENOATE (THE PRODUCT IS RAPIDLY
CC CONVERTED TO THE CORRESPONDING 12-HYDROXY COMPOUND).
CC -1- COFACTOR: IRON (CONTAINS 0.45 ATOM OF IRON PER MOLECULE).
CC -1- PATHWAY: BIOSYNTHESIS OF LEUKOTRIENES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: LEUKOCYTES, PITUITARY GLAND, LUNG AND IN
CC VERY SMALL AMOUNT IN JEJUNUM AND SPLEEN.
CC -1- TWO TYPES OF 12-LIPOXYGENASE WERE FOUND IN MAMMALIAN TISSUES:
CC ONE IN PLATELETS AND THE OTHER IN LEUKOCYTES.
CC -1- SIMILARITY: BELONGS TO THE LIPOXYGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31417; G164537; .
CC EMBL; D10621; G217680; .
CC EMBL; D10616; G217680; JOINED.
CC EMBL; D10617; G217680; JOINED.
CC EMBL; D10618; G217680; JOINED.
CC EMBL; D10619; G217680; JOINED.
CC EMBL; D10620; G217680; JOINED.
CC PIR; A35087; A35087.
CC PROSITE; PS00081; LIPOXYGENASE_2; 1.
CC PROSITE; PS00711; LIPOXYGENASE_1; 1.
CC PFAM; PF00305; lipoxigenase; 1.
CC HSP; P08170; 2SBL.
CC OXIDOREDUCTASE; DIOXYGENASE; IRON; LEUKOTRIENE BIOSYNTHESIS.
FT INIT_MET 0
FT METAL 360 360 IRON (BY SIMILARITY).
FT METAL 365 365 IRON (BY SIMILARITY).
FT METAL 540 540 IRON (BY SIMILARITY).
FT METAL 662 662 IRON (BY SIMILARITY).
SQ SEQUENCE 662 AA; 74912 MW; 304FD69D CRC32;

Query Match 88.7%; Score 55; DB 1; Length 662;
Best Local Similarity 100.0%; Pred. No. 3.19e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 133 AERKLY 139
|111111|
QY 440 AERKLY 446

RESULT 4
ID KCRU_MOUSE STANDARD; PRT; 418 AA.
AC P30275;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CREATINE KINASE, UBIQUITOUS MITOCHONDRIAL PRECURSOR (EC 2.7.3.2) (U-
DE MTCK) (MIA-CK) (ACIDIC-TYPE MITOCHONDRIAL CREATINE KINASE).
GN CKMT1.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95322032.

RESULT 3
ID LOX2_BOVIN STANDARD; PRT; 662 AA.
AC P27479;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE ARACHIDONATE 12-LIPOXYGENASE (EC 1.13.11.31) (12-LOX).
GN ALOX12.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TRACHEAL EPITHELIUM;
RX MEDLINE; 92170942.
RA DE MARZO N., SLOAN D.L., DICHARRY S., HIGHLAND E., SIGAL E.:
RT "Cloning and expression of an airway epithelial 12-lipoxygenase.";
RL AM. J. PHYSIOL. 262:L198-L207(1992).
CC -1- FUNCTION: OXYGENASE AND 14,15-LEUKOTRIENE A4 SYNTHASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ARACHIDONATE + O(2) -> (5Z,8Z,10E,14Z)-(12S)-
CC 12-HYDROPEROXYICOSA-5,8,10,14-TETRAENOATE (THE PRODUCT IS RAPIDLY
CC CONVERTED TO THE CORRESPONDING 12-HYDROXY COMPOUND).
CC -1- COFACTOR: IRON.
CC -1- PATHWAY: BIOSYNTHESIS OF LEUKOTRIENES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE LIPOXYGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81320; G162588; .
CC EMBL; M62516; G162590; .
CC EMBL; S96247; E94635; ALT SEQ.
CC PROSITE; PS00081; LIPOXYGENASE_2; 1.
CC PROSITE; PS00711; LIPOXYGENASE_1; 1.
CC PFAM; PF00305; lipoxigenase; 1.
CC OXIDOREDUCTASE; DIOXYGENASE; IRON; LEUKOTRIENE BIOSYNTHESIS.
FT INIT_MET 0
FT METAL 360 360 IRON (BY SIMILARITY).
FT METAL 365 365 IRON (BY SIMILARITY).
FT METAL 540 540 IRON (BY SIMILARITY).
FT METAL 662 662 IRON (BY SIMILARITY).
SQ SEQUENCE 662 AA; 74910 MW; B584E586 CRC32;

Query Match 88.7%; Score 55; DB 1; Length 662;
Best Local Similarity 100.0%; Pred. No. 3.19e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 133 AERKLY 139
|111111|
QY 440 AERKLY 446
```


RA STEEGHS K., PETERS W., BRUCKWILDER M., CROES H., VAN ALEWICK D.,
 RA WIERINGA B.;
 RT "Mouse ubiquitous mitochondrial creatine kinase: gene organization and
 RT consequences from inactivation in mouse embryonic stem cells.";
 RL DNA CELL BIOL. 14:539-553(1995).
 CC -1- FUNCTION: REVERSIBLY CATALYZES THE TRANSFER OF PHOSPHATE BETWEEN
 CC ATP AND VARIOUS PHOSPHOGENS (E.G. CREATINE PHOSPHATE). CREATINE
 CC KINASE ISOENZYMES PLAY A CENTRAL ROLE IN ENERGY TRANSDUCTION IN
 CC TISSUES WITH LARGE, FLUCTUATING ENERGY DEMANDS, SUCH AS SKELETAL
 CC MUSCLE, HEART, BRAIN, AND SPERMATOZOA.
 CC -1- CATALYTIC ACTIVITY: ATP + CREATINE = ADP + PHOSPHOCREATINE.
 CC -1- SUBUNIT: EXISTS AS AN OCTAMER COMPOSED OF FOUR MTCK HOMODIMERS.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE; OUTER SIDE.
 CC -1- MITOCHONDRIAL CREATINE KINASE BINDS CARDIOLIPIN.
 CC -1- SIMILARITY: THE FOUR DIFFERENT ISOZYMES OF CREATINE KINASES: B, M,
 CC AND TWO MITOCHONDRIAL FORMS; ARE HIGHLY SIMILAR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z13968; G50562; -;
 DR EMBL; Z13969; E1172459; -;
 DR PIR; S24612; S24612.
 DR MGD; MGI:99441; CKMT1.
 DR PROSITE; PS00112; GUANIDINO_KINASE; 1.
 DR PFAM; PF00217; ATP_gua_Ptrans; 1.
 DR HSP; P11009; ICRK
 KW TRANSFERASE; KINASE; MULTIGENE FAMILY; MITOCHONDRION; TRANSIT PEPTIDE.
 FT TRANSIT 1 39 MITOCHONDRION (BY SIMILARITY).
 FT CHAIN 40 418 CREATINE KINASE, UBIQUITOUS
 FT FT
 FT DOMAIN 40 64 CARDIOLIPIN-BINDING (BY SIMILARITY).
 FT ACT_SITE 317 317 BY SIMILARITY.
 FT SEQUENCE 418 AA; 47003 MW; 4BC75C76 CRC32;
 SQ
 Query Match 83.9%; Score 52; DB 1; Length 418;
 Best Local Similarity 85.7%; Pred. No. 1.53e+00;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 42 AERRRLY 48
 QY 440 AERRKLY 446
 RESULT 5
 ID YC13_ASTLO STANDARD; PRT; 456 AA.
 AC P14761;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 54.4 KD PROTEIN (ORF 456).
 GN YCF13.
 OS ASTASIA LONGA (EUGLENOPHYCEAN ALGA).
 OG CHLOROPLAST.
 OC EUKARYOTA; EUGLENOZOA; EUGLENIDA; EUGLENALES; ASTASIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CCAP 1204-17A;
 RX MEDLINE; 90251252.
 RA SIEMEISTER G., BUCHHOLZ C., HACHTEL W.;
 RT "Genes for the plastid elongation factor Tu and ribosomal protein S7
 RT and six tRNA genes on the 73 kb DNA from Astasia longa that resembles
 RT the chloroplast DNA of Euglena";
 RL MOL. GEN. GENET. 220:425-432(1990).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC STRAIN=CCAP 1204-17A;
 CC MEDLINE; 95163124.

RA GOCKEL G., HACHTEL W., BAIER S., FLISS C., HENKE M.;
 RT "Genes for components of the chloroplast translational apparatus are
 RT conserved in the reduced 73-kb plastid DNA of the nonphotosynthetic
 RT euglenoid flagellate Astasia longa.";
 RL CURR. GENET. 26:256-262(1994).
 CC -1- FUNCTION: COULD BE REQUIRED FOR GROUP III INTRON EXCISION.
 CC -1- SIMILARITY: TO GROUP II INTRON MATURASES.
 CC -1- SIMILARITY: TO A SIMILAR ORF IN E. GRACILIS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X14385; G11201; -;
 DR EMBL; X75652; G414924; -;
 DR PIR; S14924; S14924.
 DR PIR; S38613; S38613.
 DR MENDEL; S144; ASTIO; Ycf13.1.
 KW CHLOROPLAST; HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 456 AA; 54396 MW; CE2C4339 CRC32;
 Query Match 83.9%; Score 52; DB 1; Length 456;
 Best Local Similarity 71.4%; Pred. No. 1.53e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 37 DRKKLYV 43
 QY 441 ERKLYV 447
 RESULT 6
 ID HPS_HUMAN STANDARD; PRT; 700 AA.
 AC Q92902; O15502;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE HERMANSKY-PUDLAK SYNDROME PROTEIN.
 GN HPS.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANTS.
 RX MEDLINE; 97051924.
 RA OH J., BAILIN T., FUKAI K., FENG G.H., HO L., MAO J.I., FRENK E.,
 RA TAMURA N., SPRITZ R.A.;
 RT "Positional cloning of a gene for Hermansky-Pudlak syndrome, a
 RT disorder of cytoplasmic organelles.";
 RL NAT. GENET. 14:300-306(1996).
 CC [2]
 CC SEQUENCE FROM N.A., AND VARIANTS.
 RX MEDLINE; 97325811.
 RA BAILIN T., OH J., FENG G.H., FUKAI K., SPRITZ R.A.;
 RT "Organization and nucleotide sequence of the human Hermansky-Pudlak
 RT syndrome (HPS) gene";
 RL J. INVEST. DERMATOL. 108:923-927(1997).
 CC [3]
 CC SEQUENCE FROM N.A. (SHORT FORM).
 RX MEDLINE; 98239148.
 RA WILDENBERG S.C., FRYER J.P., GARDNER J.M., OETTING W.S.,
 RA BRILLIANT M.H., KING R.A.;
 RT "Identification of a novel transcript produced by the gene
 RT responsible for the Hermansky-Pudlak syndrome in Puerto Rico";
 RL J. INVEST. DERMATOL. 110:777-781(1998).
 CC -1- FUNCTION: COMPONENT OF MULTIPLE CYTOPLASMIC ORGANELLES. APPARENTLY
 CC CRUCIAL FOR THEIR NORMAL DEVELOPMENT AND FUNCTION.
 CC -1- DISEASE: DEFECTS IN HPS ARE THE CAUSE OF HERMANSKY-PUDLAK
 CC SYNDROME; AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY
 CC ALBINISM, BLEEDING, AND LYSOSOMAL STORAGE DEFECTS. THESE SYNDROMES

CC RESULT FROM DEFECTS OF DIVERSE CYTOPLASMIC ORGANELLES INCLUDING
CC MELANOSOMES, PLATELET DENSE GRANULES AND LYSOSOMES.
CC -!- DATABASE: NAME-Albinism database (ADB);
CC NOTE-HPS mutations;
CC WWW:"http://www.cbc.umn.edu/tad/".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; U65676; G1654351; -
CC DR EMBL; U96721; G2454273; -
CC DR MIM: 203300; -
CC KW ALTERNATIVE SPLICING; ALBINISM; POLYMORPHISM.
CC FT VARSPLIC 314 324 STWLEGGTTP -> EDRKAGGNNS (IN SHORT
CC FORM).
CC FT VARSPLIC 325 700 MISSING (IN SHORT FORM).
CC FT VARIANT 283 283 G -> W.
CC FT VARIANT 491 491 P -> R.
CC FT VARIANT 603 603 R -> Q.
CC FT VARIANT 630 630 V -> I.
CC FT CONFLICT 254 254 D -> H (IN REF. 3).
CC SQ SEQUENCE 700 AA; 79319 MW; 58039EE3 CRC32;

Query Match 83.9%; Score 52; DB 1; Length 700;
Best Local Similarity 100.0%; Pred. No. 1.53e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 106 RRKLYV 111
|||||
QY 442 RRKLYV 447

RESULT 7
ID PPCT_MOUSE STANDARD; PRT; 174 AA.
AC P5308;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PHOSPHATIDYLCHOLINE TRANSFER PROTEIN (PC-TP) (FRAGMENT).
GN PCTP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE; 96235171.
RA GEUTENBECK T.B.H., SMITH A.J., BORST P., WIRTZ K.W.A.;
RT "cDNA cloning and tissue-specific expression of the
RT phosphatidylcholine transfer protein gene.";
RL BIOCHEM. J. 316:49-55(1996).
CC -!- FUNCTION: CATALYZES THE TRANSFER OF PHOSPHATIDYLCHOLINE BETWEEN
CC MEMBRANES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z50024; G897786; -
CC DR MGD; MGI:107375; PCTP.
CC KW LIPID-BINDING; TRANSPORT.
CC FT NON_TER 1
CC SQ SEQUENCE 174 AA; 20540 MW; CBA17EE CRC32;

Query Match 82.3%; Score 51; DB 1; Length 174;
Best Local Similarity 71.4%; Pred. No. 2.54e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 86 DRRKIYV 92
:||||:
QY 441 ERRKLYV 447

RESULT 8
ID LOXI_HUMAN STANDARD; PRT; 661 AA.
AC P16050;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ARACHIDONATE 15-LIPOXYGENASE (EC 1.13.11.33) (ARACHIDONATE OMEGA-6
DE LIPOXYGENASE) (15-LOX).
GN ALOX15 OR LOG15.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89076270.
RA SIGAL E., CRAIK C.S., HIGHLAND E., GRUNBERGER D., COSTELLO L.L.,
RA DIXON R.A.F., NADEL J.A.;
RT "Molecular cloning and primary structure of human 15-lipoxygenase";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 157:457-464(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA KRITZIK M.R., ZIOBER A.F., SIGAL E., CONRAD D.J.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 1-15.
RX MEDLINE; 88186828.
RA SIGAL E., GRUNBERGER D., CRAIK C.S., CAUGHEY G.H., NADEL J.A.;
RT "Arachidonate 15-lipoxygenase (omega-6 lipoxygenase) from human
RT leukocytes. Purification and structural homology to other mammalian
RT lipoxygenases";
RL J. BIOL. CHEM. 263:5328-5332(1988).
RN [4]
RP SEQUENCE OF 1-30; 37-44; 156-167 AND 625-630.
RC TISSUE=EOSINOPHIL, AND LEUKOCYTE;
RX MEDLINE; 92111501.
RA IZUMI T., RADMARK O., JOERNVALL H., SAMUELSSON B.;
RT "Purification of two forms of arachidonate 15-lipoxygenase from human
RT leukocytes";
RL EUR. J. BIOCHEM. 202:1231-1238(1991).
RN [5]
RP MUTAGENESIS OF MET-417.
RX MEDLINE; 92049779.
RA SLOANE D.L., LEUNG R., CRAIK C.S., SIGAL E.;
RT "A primary determinant for lipoxygenase positional specificity";
RL NATURE 354:149-152(1991).
CC -!- CATALYTIC ACTIVITY: ARACHIDONATE + O(2) -> (5Z,8Z,11Z,13E)-(15S)-
CC 15-HYDROPEROXYICOSA-5,8,11,13-TETRAENOATE (THE PRODUCT IS RAPIDLY
CC CONVERTED TO THE CORRESPONDING 15S-HYDROXY COMPOUND).
CC -!- COFACTOR: IRON.
CC -!- PATHWAY: BIOSYNTHESIS OF LEUKOTRIENES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE LIPOXYGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M23892; G307135; -
CC DR EMBL; U68317; G1872525; -

```
DR PIR; A31349; A31349.
DR MM; 152392; -.
DR PROSITE; PS00081; LIPOXYGENASE_2; 1.
DR PROSITE; PS00711; LIPOXYGENASE_1; 1.
DR PFAM; PF00305; lipoxigenase; 1.
KW OXIDOREDUCTASE; DIOXYGENASE; IRON; LEUKOTRIENE BIOSYNTHESIS.
FT INIT_MET 0
FT METAL 359 359 IRON (BY SIMILARITY).
FT METAL 364 364 IRON (BY SIMILARITY).
FT METAL 359 539 IRON (BY SIMILARITY).
FT METAL 661 661 IRON (BY SIMILARITY).
FT MUTAGEN 417 417 M-SV: CATALYSES 15- & 12-LIPOXYGENATION.
SQ SEQUENCE 661 AA; 74673 MW; F8EB8D8E CRC32;
Query Match 82.3%; Score 51; DB 1; Length 661;
Best Local Similarity 100.0%; Pred. No. 2.54e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 133 ERRKLY 138
QY 441 ERRKLY 446
|||||
RESULT 9
ID LOX1-RABBIT STANDARD; PRT; 662 AA.
AC P12530;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ARACHIDONATE 15-LIPOXYGENASE (EC 1.13.11.33) (OMEGA-6 LIPOXYGENASE)
DE (ERYTHROID CELL-SPECIFIC 15-LIPOXYGENASE) (15-LOX).
GN ALOX15.
OS OXYTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90128296.
RA O'PREY J., CHESTER J., THIELE B.J., JANETZKI S., PREHN S., FLEMING J.,
RA HARRISON P.R.;
RT "The promoter structure and complete sequence of the gene encoding
RT the rabbit erythroid cell-specific 15-lipoxygenase.";
RL GENE 84.493-499(1989).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-30.
RX MEDLINE; 89378774.
RA FLEMING J., THIELE B.J., CHESTER J., O'PREY J., JANETZKI S.,
RA ATKEN A., ANTON I.A., RAPOPORT S.M., HARRISON P.R.;
RT "The complete sequence of the rabbit erythroid cell-specific 15-
RT lipoxygenase mRNA: comparison of the predicted amino acid sequence of
RT the erythrocyte lipoxygenase with other lipoxygenases.";
RL GENE 79:181-188(1989).
RN [3]
RP SEQUENCE OF 1-30 FROM N.A.
RX MEDLINE; 88112854.
RA THIELE B.J., FLEMING J., KASTURI K., O'PREY J., BLACK E., CHESTER J.,
RA RAPOPORT S.M., HARRISON P.R.;
RT "Cloning of a rabbit erythroid-cell-specific lipoxygenase mRNA.";
RL GENE 57:111-119(1987).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE; 98069848.
RA GILLMOR S.A., VILLASENOR A., FLETTERICK R., SIGAL E., BROWNER M.F.;
RT "The structure of mammalian 15-lipoxygenase reveals similarity to the
RT lipases and the determinants of substrate specificity.";
RL NAT. STRUCT. BIOL. 4:1003-1009(1997).
CC -1- CATALYTIC ACTIVITY: ARACHIDONATE + O(2) -> (5Z,8Z,11Z,13E)-(15S)-
CC 15-HYDROPEROXYICOSA-5,8,11,13-TETRAENOATE (THE PRODUCT IS RAPIDLY
CC CONVERTED TO THE CORRESPONDING 15S-HYDROXY COMPOUND).
CC -1- COFACTOR: IRON.
CC -1- PATHWAY: BIOSYNTHESIS OF LEUKOTRIENES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE LIPOXYGENASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M22617; G536862; -.
DR EMBL; M33291; G164732; -.
DR EMBL; M27214; G164907; -.
DR PIR; JQ0018; JQ0018.
DR PIR; A27327; A27327.
DR PIR; JC1513; JC1513.
DR PROSITE; PS00081; LIPOXYGENASE_2; 1.
DR PROSITE; PS00711; LIPOXYGENASE_1; 1.
DR PFAM; PF00305; lipoxigenase; 1.
DR PDB; 1LOX; O2-DEC-98.
KW OXIDOREDUCTASE; DIOXYGENASE; IRON; LEUKOTRIENE BIOSYNTHESIS;
KW 3D-STRUCTURE.
FT INIT_MET 0
FT METAL 360 360 IRON (BY SIMILARITY).
FT METAL 365 365 IRON (BY SIMILARITY).
FT METAL 540 540 IRON (BY SIMILARITY).
FT METAL 662 662 IRON (BY SIMILARITY).
FT CONFLICT 112 112 T -> I (IN REF. 2).
FT CONFLICT 189 189 N -> D (IN REF. 2).
FT CONFLICT 193 193 I -> V (IN REF. 2).
SQ SEQUENCE 662 AA; 75178 MW; 0642DB4D CRC32;
Query Match 82.3%; Score 51; DB 1; Length 662;
Best Local Similarity 100.0%; Pred. No. 2.54e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 134 ERRKLY 139
QY 441 ERRKLY 446
|||||
RESULT 10
ID RNPI-YEAST STANDARD; PRT; 249 AA.
AC P22385;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE RIBONUCLEOPROTEIN-1.
GN RNPI OR YLL046C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94188141.
RA CUSICK M.E.;
RT "RNPI, a new ribonucleoprotein gene of the yeast Saccharomyces
RT cerevisiae.";
RL NUCLEIC ACIDS RES. 22:869-877(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA WEDLER H., WEDLER E., SCHARFE M., WAMBUTT R.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RNP).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M88608; G460047; -.
-----
```

```
DR EMBL: 273152; E245473; -.
DR PIR: S27456; S27456.
DR SGD: L0001654; RNPI.
DR PROSITE: PS00030; RNP_1; FALSE_NEG.
DR PFAM: PF00076; itm: 1.
KW RNA-BINDING; REPEAT.
FT DOMAIN 37 42 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 77 84 RNA-BINDING (RNP1) (BY SIMILARITY).
FT DOMAIN 142 147 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 181 188 RNA-BINDING (RNP1) (BY SIMILARITY).
SQ SEQUENCE 249 AA; 28861 MW; 89B85504 CRC32;

Query Match      80.6%; Score 50; DB 1; Length 249;
Best Local Similarity 71.4%; Pred. No. 4.19e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 33 DRETLV 39
   :|||
OY 441 ERRLV 447

RESULT 11
ID KCRU HUMAN STANDARD; PRT: 417 AA.
AC P12532;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CREATINE KINASE, UBIQUITOUS MITOCHONDRIAL PRECURSOR (EC 2.7.3.2) (U-
DE MTCK) (MIA-CK) (ACIDIC-TYPE MITOCHONDRIAL CREATINE KINASE).
GN CKMT1 OR CKMT.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89123390.
RA HAAS R.C., KORENFELD C., ZHANG Z., PERRYMAN B., ROMAN D.,
RA STRAUSS A.W.;
RT "Isolation and characterization of the gene and cDNA encoding human
RT mitochondrial creatine kinase.";
RL J. BIOL. CHEM. 264:2890-2897(1989).
CC -!- FUNCTION: REVERSIBLY CATALYZES THE TRANSFER OF PHOSPHATE BETWEEN
CC ATP AND VARIOUS PHOSPHOGENS (E.G. CREATINE PHOSPHATE). CREATINE
CC KINASE ISOENZYMES PLAY A CENTRAL ROLE IN ENERGY TRANSDUCTION IN
CC TISSUES WITH LARGE, FLUCTUATING ENERGY DEMANDS, SUCH AS SKELETAL
CC MUSCLE, HEART, BRAIN, AND SPERMATOZOA.
CC -!- CATALYTIC ACTIVITY: ATP + CREATINE -> ADP + PHOSPHOCREATINE.
CC -!- SUBUNIT: EXISTS AS AN OCTAMER COMPOSED OF FOUR MTCK HOMODIMERS.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE; OUTER SIDE.
CC -!- MITOCHONDRIAL CREATINE KINASE BINDS CARDIOLIPIN.
CC -!- SIMILARITY: THE FOUR DIFFERENT ISOZYMES OF CREATINE KINASES: B, M,
CC AND TWO MITOCHONDRIAL FORMS: ARE HIGHLY SIMILAR.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: J04469; G180590; -.
DR PIR: A30789; A30789.
DR PIR: A31431; A31431.
DR MIM: 123290; -.
DR PROSITE: PS00112; GUANIDO_KINASE; 1.
DR PFAM: PF00217; ATP_gua_Pfams; 1.
DR HSP: F11009; 1CRK.
KW TRANSFERASE; KINASE; MULTIGENE FAMILY; MITOCHONDRION; TRANSIT PEPTIDE.
FT TRANSIT 1 39 MITOCHONDRION.
FT CHAIN 40 417 CREATINE KINASE, UBIQUITOUS
FT DOMAIN 40 54 MITOCHONDRIAL.
FT CARDIOLIPIN-BINDING (BY SIMILARITY).
```

```
FT ACT_SITE 316 316 BY SIMILARITY.
SQ SEQUENCE 417 AA; 47036 MW; E82E7844 CRC32;

Query Match      80.6%; Score 50; DB 1; Length 417;
Best Local Similarity 71.4%; Pred. No. 4.19e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 41 SERRLY 47
   :|||:|
OY 440 AERRLV 446

RESULT 12
ID GBA4_USTMA STANDARD; PRT: 580 AA.
AC P87035;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GUANINE NUCLEOTIDE-BINDING PROTEIN ALPHA-4 SUBUNIT.
GN GPA4.
OS USTILAGO MAYDIS (SMUT FUNGUS).
OC EUKARYOTA; FUNGI; BASIDIOMYCOTA; USTILAGINOMYCETES;
OC USTILAGINOMYCETIDAE; USTILAGINALES; USTILAGINACEAE; USTILAGO.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FBI;
RX MEDLINE: 97299870.
RA REGENFELDER E., SPELLIG T., HARTMANN A., LAUENSTEIN S., BOLKER M.,
RA KAHMANN R.;
RT "G proteins in Ustilago maydis: transmission of multiple signals?";
RL EMBO J. 16:1934-1942(1997).
CC -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC INVOLVED AS MODULATORS OR TRANSDUCERS IN VARIOUS TRANSMEMBRANE
CC SIGNALING SYSTEMS.
CC -!- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC THE ALPHA CHAIN CONTAINS THE GUANINE NUCLEOTIDE BINDING SITE.
CC -!- SIMILARITY: BELONGS TO THE G-ALPHA FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U85778; G1930025; -.
DR PFAM: PF00503; G-alpha; 1.
DR HSP: P10824; 1AS3.
KW GTP-BINDING; TRANSDUCER; MULTIGENE FAMILY.
FT NP_BIND 101 108 GTP (BY SIMILARITY).
FT NP_BIND 415 418 GTP (BY SIMILARITY).
FT NP_BIND 484 487 GTP (BY SIMILARITY).
SQ SEQUENCE 580 AA; 65198 MW; 10EF3455 CRC32;

Query Match      80.6%; Score 50; DB 1; Length 580;
Best Local Similarity 62.5%; Pred. No. 4.19e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 539 TSRRRLV 546
   :||:|
OY 440 AERRLV 447

RESULT 13
ID RAD1_YEAST STANDARD; PRT: 1100 AA.
AC P06777;
DT 01-JAN-1988 (REL. 05, CREATED)
DT 01-JAN-1988 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DNA REPAIR PROTEIN RAD1.
GN RAD1 OR YPO22W.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
```

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 8712766.
RA REYNOLDS P., PRAKASH L., PRAKASH S.;
RT "Nucleotide sequence and functional analysis of the RAD1 gene of
Saccharomyces cerevisiae."
RL MOL. CELL. BIOL. 7:1012-1020(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85061207.
RA YANG E., FRIEDBERG E.C.;
RT "Molecular cloning and nucleotide sequence analysis of the
Saccharomyces cerevisiae RAD1 gene."
RL MOL. CELL. BIOL. 4:2161-2169(1984).
RN [3]
RP FUNCTION.
RX MEDLINE; 93241303.
RA TOMKINSON A.E., BARDWELL A.J., BARDWELL L., TAPPE N.J.,
RA FRIEDBERG E.C.;
RT "Yeast DNA repair and recombination proteins Rad1 and Rad10
constitute a single-stranded-DNA endonuclease."
RL NATURE 362:860-862(1993).
CC -!- FUNCTION: INVOLVED IN NUCLEOTIDE EXCISION REPAIR OF DNA DAMAGED
WITH UV LIGHT, BULKY ADDUCTS, OR CROSS-LINKING AGENTS. ALONG WITH
RAD10 FORMS AN ENDONUCLEASE THAT SPECIFICALLY DEGRADES SINGLE-
STRANDED DNA.
CC -!- SUBUNIT: HETERODIMER COMPOSED OF RAD1 AND RAD10.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE XPF/ERCC4/RAD1/RAD16 FAMILY.
CC -!- CAUTION: REF 2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE C-TERMINUS
DUE TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U36624; G1039455; .
DR EMBL; M15435; G172325; .
DR EMBL; K02070; G172318; ALT_FRAME.
DR PIR; A26129; DDBYD1.
DR SGD; L0001555; RAD1.
KW DNA REPAIR; DNA-BINDING; NUCLEAR PROTEIN; HYDROLASE; NUCLEASE;
FT DOMAIN 1 110 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 516 576 ARG/LYS-RICH (BASIC).
FT DOMAIN 595 689 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1041 1100 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 223 223 D -> N (IN REF. 2).
FT CONFLICT 883 883 C -> Y (IN REF. 2).
FT CONFLICT 886 886 E -> I (IN REF. 2).
FT CONFLICT 912 912 M -> K (IN REF. 2).
FT SEQUENCE 1100 AA; 126369 MW; 129AD61A CRC32;
Query Match 80.6%; Score 50; DB 1; Length 1100;
Best Local Similarity 83.3%; Pred. No. 4.19e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 203 RRKLYI 208
QY 442 RRKLYV 447
RESULT 14
ID FLIH_CAEEL STANDARD; PRT; 1257 AA.
AC P34268; P90739; Q17331.
DT 01-FEB-1994 (REL. 28, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE FLIGHTLESS-I PROTEIN HOMOLOG.
GN B0523.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94068608.
RA CAMPBELL H.D., SCHIMANSKY T., CLAUDIANOS C., OZSARAC N.,
RA KASPRZAK A.B., COTSELL J.N., YOUNG I.G., DE COUET H.G., MIKLOS G.L.G.;
RT "The Drosophila melanogaster flightless-I gene involved in
gastrulation and muscle degeneration encodes gelsolin-like and
leucine-rich repeat domains and is conserved in Caenorhabditis
elegans and humans."
RL PROC. NATL. ACAD. SCI. U.S.A. 90:11386-11390(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSIER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SINS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOULDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans."
RL NATURE 368:32-38(1994).
RN [3]
RP REVISIONS.
RA WATERSTON R.;
RL SUBMITTED (JUN-1996) TO THE SWISS-PROT DATA BANK.
CC -!- FUNCTION: MAY PLAY A KEY ROLE IN EMBRYONIC CELLULARIZATION BY
INTERACTING WITH BOTH THE CYTOSKELETON AND OTHER CELLULAR
COMPONENTS (BY SIMILARITY).
CC -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
MANY PROTEINS. NUMBER IN THIS PROTEIN: 16.
CC -!- SIMILARITY: BELONGS TO THE VILLIN/GELSOLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U01183; G440175; .
DR EMBL; L07143; G1695248; .
DR PIR; S27783; S27783.
DR WORMPEP; B0523.5; CE06721.
DR PFAM; PF00560; LRR; 8.
DR PFAM; PF00626; Gelsolin; 4.
DR HSP; P10733; 1SVQ;
KW DEVELOPMENTAL PROTEIN; REPEAT; LEUCINE-REPEAT.
FT DOMAIN 1 374 LEUCINE-RICH REPEATS.
FT REPEAT 1 21 LRR 1.
FT REPEAT 22 44 LRR 2.
FT REPEAT 45 67 LRR 3.
FT REPEAT 68 92 LRR 4.
FT REPEAT 93 115 LRR 5.
FT REPEAT 116 139 LRR 6.
FT REPEAT 140 162 LRR 7.
FT REPEAT 163 186 LRR 8.
FT REPEAT 187 211 LRR 9.
FT REPEAT 212 234 LRR 10.
FT REPEAT 235 257 LRR 11.

```
FT REPEAT      258      280      LRR 12.
FT REPEAT      281      305      LRR 13.
FT REPEAT      306      328      LRR 14.
FT REPEAT      329      351      LRR 15.
FT REPEAT      352      374      LRR 16.
FT DOMAIN      501      592      GELSOLIN-LIKE.
FT CONFLICT    154      154      F -> S (IN REF. 3).
SQ SEQUENCE    1257 AA; 144814 MW; 1B0888A1 CRC32;

Query Match      80.6%; Score 50; DB 1; Length 1257;
Best Local Similarity 85.7%; Pred. No. 4.19e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 738 AERKKLY 744
   |||:|
Qy 440 AERRKLY 446

RESULT 15
ID YKX2_YEAST STANDARD; PRT; 101 AA.
AC P34249;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 11.4 KD PROTEIN IN APE1/LAP4-CWPI INTERGENIC REGION.
GN YKL102C OR YKL454.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETACEAE;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE; 94152173.
RA CHERET G., PALLIER C., VALENS M., DAIGNAN-FORNIER B., FUKUHARA H.,
RA BOLOTIN-FUKUHARA M., SOR F.;
RT "The DNA sequence analysis of the HAP4-LAP4 region on chromosome XI
RT of Saccharomyces cerevisiae suggests the presence of a second
RT aspartate aminotransferase gene in yeast.";
RL YEAST 9:1259-1265(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X71113; G431214; -.
DR EMBL; Z28102; G486171; -.
PIR: S37929; S37929.
DR PIR: S39102; S39102.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 101 AA; 11423 MW; B7E94D38 CRC32;

Query Match      79.0%; Score 49; DB 1; Length 101;
Best Local Similarity 77.8%; Pred. No. 6.85e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 34 SERRKVLV 42
   :| | | |
Qy 440 AERRK-LYV 447
```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

```
Run On: Thu Jul 8 18:55:25 1999; Maspar time 4.92 Seconds
88.709 Million cell updates/sec
```

Tabular output not generated.

```
>US-09-041-236-2
Title:
Description: (440-447) from US09041236.pep (24 of 45)
Perfect Score: 62
Sequence: 1 AERRKLYV 8
```

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

sp:treml9

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_ordanelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 23.771; Variance 26.691; scale 0.891

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match	%					
1	62	100.0		666	4	075326	SEMAPHORIN L.	9.81e+03
2	55	88.7		79	4	Q12938	12-LIPOXYGENASE (FRAGM	4.79e+01
3	52	83.9		359	11	G62941	ZINC TRANSPORTER ZNT-2	2.32e+00
4	52	83.9		676	4	015402	HERMANSKY-PUDLAK SYNDR	2.32e+00
5	52	83.9		934	1	073954	DNA TOPOISOMERASE I (F	2.32e+00
6	52	83.9		1729	5	Q25734	PREMP1 VARIANT 2 OF ST	2.32e+00
7	52	83.9		2135	5	G61077	PREMP1 VARIANT 2 OF ST	2.32e+00
8	52	83.9		2924	5	Q25733	VARIANT-SPECIFIC SURFA	2.32e+00
9	52	83.9		3026	5	G26030	PREMP1 VARIANT 1 OF ST	2.32e+00
10	51	82.3		224	5	Q21677	VARIANT SURFACE PROTEI	2.32e+00
11	51	82.3		663	6	019043	COSMID R03H10.	3.87e+00
12	50	80.6		345	5	Q22305	12-LIPOXYGENASE.	3.87e+00
13	50	80.6		510	2	G05044	T07D4.1 PROTEIN.	6.41e+00
14	50	80.6		2706	5	Q15870	ORF510.	6.41e+00
15	49	79.0		146	2	Q48381	PREMP1 (FRAGMENT).	1.06e+01
16	49	79.0		188	13	Q73682	MOAI PROTEIN.	1.06e+01
17	49	79.0		201	5	Q22508	VASCULAR ENDOTHELIAL G	1.06e+01
18	49	79.0		602	3	Q93932	CODED FOR BY C. ELEGAN	1.06e+01
19	49	79.0		684	4	Q00307	ORF_YOR324C.	1.06e+01
20	49	79.0		2182	5	Q26034	WWP1 (FRAGMENT).	1.06e+01
							VARIANT-SPECIFIC SURFA	1.06e+01

21	49	79.0	2664	5	Q26033	VARIANT-SPECIFIC SURFA	1.06e+01
22	49	79.0	3078	5	Q26031	VARIANT-SPECIFIC SURFA	1.06e+01
23	48	77.4	82	2	P71325	COLI COLD-CA23 PLASMID	1.72e+01
24	48	77.4	87	2	O52990	COLICIN D-157 IMMUNITY	1.72e+01
25	48	77.4	385	1	O50119	385AA LONG HYPOTHETICA	1.72e+01
26	48	77.4	499	14	Q87029	UNIDENTIFIED GENES, TH	1.72e+01
27	48	77.4	785	14	P84451	HYPERV SIMPLEX VIRUS T	1.72e+01
28	48	77.4	885	5	Q24801	PERUVATE PHOSPHATE DIK	1.72e+01
29	48	77.4	1456	14	O85245	PUTATIVE PEPTIDASE COM	1.72e+01
30	48	77.4	1456	14	O85198	165KDA PROTEIN	1.72e+01
31	48	77.4	4096	4	Q13227	DNA DEPENDENT PROTEIN	1.72e+01
32	48	77.4	4127	4	P78527	DNA-DEPENDENT PROTEIN	1.72e+01
33	47	75.8	110	1	O58149	110AA LONG HYPOTHETICA	2.79e+01
34	47	75.8	228	2	O66557	TRANSCRIPTIONAL REGULA	2.79e+01
35	47	75.8	317	2	O56420	ORF 1A.	2.79e+01
36	47	75.8	317	2	O56421	ORF 1B.	2.79e+01
37	47	75.8	591	10	O82178	74C15.20 PROTEIN.	2.79e+01
38	47	75.8	688	5	Q26557	GYNECOPHORAL CANAL PRO	2.79e+01
39	47	75.8	789	5	O16781	SIMILARITY TO MULTIPLE	2.79e+01
40	47	75.8	804	1	Q29777	CATION-TRANSPORTING AT	2.79e+01
41	47	75.8	870	4	O00308	WWP2	2.79e+01
42	47	75.8	1045	13	Q91292	INTEGRIN.	2.79e+01
43	47	75.8	1908	5	Q17971	G14C10.5 PROTEIN.	2.79e+01
44	47	75.8	2228	5	O60991	ERYTHROCYTE MEMBRANE P	2.79e+01
45	47	75.8	2647	5	P90580	FCR3-VART11-1 PROTEIN	2.79e+01

ALIGNMENTS

RESULT	1	
ID	O75326	PRELIMINARY; PRT; 666 AA.
AC	O75326;	
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE	SENAPHORIN L.	
GN	SENAL	
OS	HOMO SAPIENS (HUMAN).	
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;	
OC	CATARRHINI; HOMINIDAE; HOMO.	

Query Match 100.0%; Score 62; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 9.81e-03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 472 AERRKLYV 479
 |||||
 Ov 440 AERRKLYV 447

RESULT 2
ID Q12938
PRELIMINARY;
PRT; 79 AA.

Q12936:
AC 01-NOV-1996 (TREMBLREL. 01, CREATED)
CT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 12-LIPOXYGENASE (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
CATARRHINI; HOMINIDAE; HOMO.
[1]
RN
RP SEQUENCE FROM N.A.
RA NADLER J., ROSSI J.J., YAMAMOTO S., YOSHIMOTO T., BEN-EZRA J.,
RA VALENTE G., NATARAJAN R., GU J.L.;
RA

RL SUBMITTED (JUN-1992) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; M95923; G553159; -.
DR PFAM; PF00305; lipoxigenase; 1.
FT NON_TER 1
SQ SEQUENCE 79 AA; 9448 MW; BBD2EA2A CRC32;

Query Match 88.7%; Score 55; DB 4; Length 79;
Best Local Similarity 100.0%; Pred. No. 4.79e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 68 AERRKLY 74
|||||
QY 440 AERRKLY 446

RESULT 3
ID Q62941 PRELIMINARY; PRT; 359 AA.
AC Q62941;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE ZINC TRANSPORTER ZNT-2.
GN ZNT-2.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE DAWLEY;
RX MEDLINE; 96203098.
RA PALMITER R.D., COLE T.B., FINDLEY S.D.;
RT "ZNT-2, a mammalian protein that confers resistance to zinc by
RT facilitating vesicular sequestration.";
RL EMBO J. 15:1784-1791(1996).
DR EMBL; U50927; G1256378; -.
SQ SEQUENCE 359 AA; 39276 MW; FEEB4EE CRC32;

Query Match 83.9%; Score 52; DB 11; Length 359;
Best Local Similarity 100.0%; Pred. No. 2.32e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 54 RRKLYV 59
|||||
QY 442 RRKLYV 447

RESULT 4
ID O15402 PRELIMINARY; PRT; 676 AA.
AC O15402;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE HERMANSKY-PUDLAK SYNDROME PROTEIN.
GN HPS.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97325811.
RA BAILIN T., OH J., FENG G.H., FUKAI K., SPRITZ R.A.;
RT "Organization and nucleotide sequence of the human Hermansky-Pudlak
RT syndrome (HPS) gene.";
RL J. INVEST. DERMATOL. 108:923-927(1997).
DR EMBL; U79136; G2408213; -.
DR EMBL; U79123; G2408213; JOINED.
DR EMBL; U79124; G2408213; JOINED.
DR EMBL; U79125; G2408213; JOINED.
DR EMBL; U79126; G2408213; JOINED.
DR EMBL; U79127; G2408213; JOINED.
DR EMBL; U79128; G2408213; JOINED.
DR EMBL; U79129; G2408213; JOINED.
DR EMBL; U79130; G2408213; JOINED.

DR EMBL; U79131; G2408213; JOINED.
DR EMBL; U79132; G2408213; JOINED.
DR EMBL; U79133; G2408213; JOINED.
DR EMBL; U79134; G2408213; JOINED.
DR EMBL; U79135; G2408213; JOINED.
SQ SEQUENCE 676 AA; 76520 MW; A3886DA6 CRC32;

Query Match 83.9%; Score 52; DB 4; Length 676;
Best Local Similarity 100.0%; Pred. No. 2.32e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 106 RRKLYV 111
|||||
QY 442 RRKLYV 447

RESULT 5
ID O73954 PRELIMINARY; PRT; 934 AA.
AC O73954;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE DNA TOPOISOMERASE I (FRAGMENT).
GN TOPA.
OS PYROCOCCLUS FURIOSUS.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA CHUTE I.C., HU Z., LIU X.-Q.;
RT "A topa intein in Pyrococcus furiosus and its relatedness to the
RT r-gyr intein of Methanococcus jannaschii.";
RL GENE 0:0-0(1998).
DR EMBL; AF042825; G3309559; -.
KW ISOMERASE.
FT NON_TER 1
SQ SEQUENCE 934 AA; 107008 MW; A847FB7D CRC32;

Query Match 83.9%; Score 52; DB 1; Length 934;
Best Local Similarity 100.0%; Pred. No. 2.32e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 493 RRKLYV 498
|||||
QY 442 RRKLYV 447

RESULT 6
ID Q25734 PRELIMINARY; PRT; 1729 AA.
AC Q25734;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE PFEMP1 VARIANT 2 OF STRAIN MC (FRAGMENT).
GN MCVAR-2 PFEMP1.
OS PLASMODIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MALAYAN CAMP (MC132 K+C+R+);
RX MEDLINE; 95330812.
RA BARUCH D.I., PASLOSKE B.L., SINGH H.B., BI X., MA X.C., FELDMAN M.,
RA TARASCH T.F., HOWARD R.J.;
RT "Cloning the p. falciparum gene encoding PfEMP1, a malarial variant
RT antigen and adherence receptor on the surface of parasitized human
RT erythrocytes.";
RL CELL 82:77-87(1995).
DR EMBL; U27339; G914921; -.
FT NON_TER 1729 1729
SQ SEQUENCE 1729 AA; 195156 MW; 2D71EE95 CRC32;

Query Match 83.9%; Score 52; DB 5; Length 1729;
Best Local Similarity 100.0%; Pred. No. 2.32e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 895 RRKLYV 900
|||||
Qy 442 RRKLYV 447

RESULT 7
ID O61077 PRELIMINARY; PRT: 2135 AA.
AC O61077
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE VARIANT-SPECIFIC SURFACE PROTEIN (FRAGMENT).
GN VARP17.
OS PLASMIDIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA VOSS T.S., FELGER I., WEISS N., BECK H.-P.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF050740; G2944095; -;
FT NON_TER 2135 2135
SQ SEQUENCE 2135 AA; 242712 MW; 2DE4C2B5 CRC32;

Query Match 83.9%; Score 52; DB 5; Length 2135;
Best Local Similarity 100.0%; Pred. No. 2.32e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1448 RRKLYV 1453
|||||
Qy 442 RRKLYV 447

RESULT 8
ID Q25733 PRELIMINARY; PRT: 2924 AA.
AC Q25733;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PFEMP1 VARIANT 1 OF STRAIN MC.
GN MCVAR-1 PFEMP1.
OS PLASMIDIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-MALAYAN CAMP (MC132 K+C+R+);
RX MEDLINE: 95330812.
RA BARUCH D.I., PASLOSKE B.L., SINGH H.B., BI X., MA X.C., FELDMAN M.,
RA TARASCHI T.F., HOWARD R.J.;
RT "Cloning the P. falciparum gene encoding PfEMP1, a malarial variant
antigen and adherence receptor on the surface of parasitized human
erythrocytes.";
RL CELL 82:77-87(1995).
DR EMBL: U27338; G914919; -;
FT VARIANT 104 104 S -> T.
FT VARIANT 105 105 E -> K.
FT VARIANT 139 139 D -> G.
FT VARIANT 362 362 K -> E.
FT VARIANT 503 503 K -> N.
FT VARIANT 581 581 S -> T.
FT VARIANT 593 593 M -> V.
FT VARIANT 617 617 I -> K.
FT VARIANT 631 631 V -> G.
FT VARIANT 658 658 D -> E.
FT VARIANT 722 722 K -> N.
SQ SEQUENCE 2924 AA; 335862 MW; CC018C62 CRC32;

Query Match 83.9%; Score 52; DB 5; Length 2924;
Best Local Similarity 100.0%; Pred. No. 2.32e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 903 RRKLYV 908
|||||

Qy 442 RRKLYV 447
RESULT 9
ID Q26030 PRELIMINARY; PRT: 3026 AA.
AC Q26030;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE VARIANT SURFACE PROTEIN (FRAGMENT).
GN VAR.
OS PLASMIDIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMIDIUM.
RN [1]
RP SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN-IT 4/25/5;
RA SMITH J.D., CHITNIS C.E., CRAIG A.G., ROBERTS D.J.,
RA HUDSON-TAYLOR D.E., PETERSON D.S., PINCHES R., NEWBOLD C.I.,
RA MILLER L.H.;
RT "Switches in the expression of Plasmodium falciparum var genes
correlate with changes in antigenic and cytoadherent phenotypes of
infected erythrocytes.";
RL CELL 0:0-0(1995).
RN [2]
RP SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN-IT 4/25/5;
RX MEDLINE: 95330813.
RA SU X.Z., HEATWOLE V.M., WERTHEIMER S.P., GUINET F., HERRFELDT J.A.,
RA PETERSON D.S., RAVETCH J.A., WELLEMS T.E.;
RT "The large diverse gene family var encodes proteins involved in
cytoadherence and antigenic variation of Plasmodium
falciparum-infected erythrocytes.";
RL CELL 82:89-100(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA SMITH J.D., KYES S.A., CRAIG A.G., FAGAN T., HUDSON-TAYLOR D.,
RA MILLER L.H., BARUCH D.I., NEWBOLD C.I.;
RT "Heterologous expression identifies an adhesive domain in A4
Plasmodium falciparum erythrocyte membrane protein-1.";
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: L42244; G3540145; -;
FT NON_TER 3026 3026
SQ SEQUENCE 3026 AA; 348124 MW; D614D7F8 CRC32;

Query Match 83.9%; Score 52; DB 5; Length 3026;
Best Local Similarity 100.0%; Pred. No. 2.32e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1285 RRKLYV 1290
|||||
Qy 442 RRKLYV 447

RESULT 10
ID Q21677 PRELIMINARY; PRT: 224 AA.
AC Q21677;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COSMID R03H10.
GN R03H10.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WILCOX L.;
 RN [3]
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U29382; G868235; -;
 SQ SEQUENCE 224 AA; 24895 MW; 2BA30F2E CRC32;

Query Match 82.3%; Score 51; DB 5; Length 224;
 Best Local Similarity 62.5%; Pred. No. 3.87e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 123 GFKRLYV 130
 :||:||||
 Qy 440 AERRKLYV 447

RESULT 11
 ID O19043 PRELIMINARY; PRT; 663 AA.
 AC O19043;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE 12-LIPOXYGENASE.
 OS ORYCTOLAGUS CUNICULUS (RABBIT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BERGER M., SCHWARZ K., THIELE H., REIMANN I., HUTH A., BORNGRAEBER S.,
 RA KUEHN H., THIELE B.J.;
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; 297654; E329364; -;
 DR PROSITE; PS00081; LIPOXYGENASE_2; 1.
 DR PROSITE; PS00711; LIPOXYGENASE_1; 1.
 DR PFAM; PF00305; lipoxigenase; 1.
 SQ SEQUENCE 663 AA; 75233 MW; 52A3B482 CRC32;

Query Match 82.3%; Score 51; DB 6; Length 663;
 Best Local Similarity 100.0%; Pred. No. 3.87e+00;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 135 ERRLY 140
 :|||||
 Qy 441 ERRLY 446

RESULT 12
 ID Q22305 PRELIMINARY; PRT; 345 AA.
 AC Q22305;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE T07D4.1 PROTEIN.
 GN T07D4.1
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]

RP SEQUENCE FROM N.A.
 RA SMYE R.;
 RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL NATURE 368:32-38(1994).
 DR EMBL; Z50071; E134906; -;
 SQ SEQUENCE 345 AA; 39371 MW; 49D13A9A CRC32;

Query Match 80.6%; Score 50; DB 5; Length 345;
 Best Local Similarity 83.3%; Pred. No. 6.41e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 295 RRKLYI 300
 :||||:
 Qy 442 RRKLYV 447

RESULT 13
 ID O50544 PRELIMINARY; PRT; 510 AA.
 AC O50544;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE ORF510.
 OS SYNECHOCOCCUS SP. (STRAIN PCC 7942).
 OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC 7942;
 RA PHUNG L.T., HASELKORN R.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U59234; G2661139; -;
 SQ SEQUENCE 510 AA; 56679 MW; F381E30A CRC32;

Query Match 80.6%; Score 50; DB 2; Length 510;
 Best Local Similarity 62.5%; Pred. No. 6.41e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 450 GERRLYL 457
 :||||:
 Qy 440 AERRKLYV 447

RESULT 14
 ID O15870 PRELIMINARY; PRT; 2706 AA.
 AC O15870;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE PEEMF1 (FRAGMENT).
 GN R29R+VAR1.
 OS PLASMODIUM FALCIPARUM.
 OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IT 4/25/5;
 RX MEDLINE; 97373957.
 RA ROWE J.A., MOULDS J.M., NEWBOLD C.I., MILLER L.H.;
 RT "P. falciparum rosetting mediated by a parasite-variant erythrocyte

RT membrane protein and complement-receptor 1.";
 RL NATURE 388:292-295(1997).
 DR EMBL; Y13402; E321408; .
 DR EMBL; Y13403; E321408; JOINED.
 FT NON_TER 2706 2706
 SQ SEQUENCE 2706 AA; 308164 MW; 3ALCBED0 CRC32;

Query Match 80.6%; Score 50; DB 5; Length 2706;
 Best Local Similarity 83.3%; Pred. No. 6.41e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1671 RKLYI 1676
 |||||
 Qy 442 RKLYV 447

RESULT 15

ID Q48381 PRELIMINARY; PRT; 146 AA.
 AC Q48381;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE MOAI PROTEIN.
 GN MOAI.
 OS KLEBSIELLA AEROGENES.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC KLEBSIELLA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=w70;
 RA SUGINO H.;
 RT "Cloning and characterization of the moai gene from Klebsiella
 aerogenes.";
 RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
 DR EMBL; D69524; D1010436; .
 DR PFAM; PF01047; Mar; 1.
 SQ SEQUENCE 146 AA; 16712 MW; 53345E60 CRC32;

Query Match 79.0%; Score 49; DB 2; Length 146;
 Best Local Similarity 71.4%; Pred. No. 1.06e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 83 DOKLYV 89
 :|||||
 Qy 441 ERKLYV 447

Search completed: Thu Jul 8 18:55:50 1999
 Job time : 25 secs.

THIS PAGE BLANK (USPTO)

W E S R E D (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 18:57:18 1999; MasPar time 3.29 Seconds
Tabular output not generated. 68.748 Million cell updates/sec

Title: >US-09-041-236-2
Description: (475-482) from US09041236.pap (25 of 45)
Perfect Score: 79
Sequence: 1 RDPYCGWD 8

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 24.241; Variance 31.212; scale 0.777

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	56	70.9	590	1 NU5M_TRYBB	NADH-UBIQUINONE OXIDOR	6.96e-01
2	56	70.9	1110	1 CYGX_RAT	OLFACTORY GUANYLYL CYC	6.96e-01
3	56	70.9	1157	1 PEX1_PICPA	PEROXISOME BIOSYNTHESI	6.96e-01
4	56	70.9	3163	1 POLG_TUMVQ	GENOME POLYPROTEIN [CO	6.96e-01
5	56	70.9	3164	1 POLG_TUMVJ	GENOME POLYPROTEIN [CO	6.96e-01
6	55	69.6	2255	1 RPLP_SVSWR	RNA POLYMERASE BETA SU	1.11e+00
7	55	69.6	2255	1 RPLP_SVSVS	RNA POLYMERASE BETA SU	1.11e+00
8	54	68.4	286	1 AC8_STRFR	AMINOGLYCOSIDE N3'-ACE	1.76e+00
9	54	68.4	417	1 AGP_PRORE	GLUCOSE-1-PHOSPHATASE	1.76e+00
10	53	67.1	225	1 UL04_HSVB7	GENE 58 PROTEIN	2.78e+00
11	53	67.1	353	1 RPO4_EVSP	RNA REPLICATION PROTEI	2.78e+00
12	53	67.1	471	1 GLPD_ECOLI	GLUTAMATE SYNTHASE [NA	2.78e+00
13	53	67.1	1291	1 YCB1_CAEEL	HYPOTHETICAL 143.1 KD	2.78e+00
14	52	65.8	113	1 HEMT_THEDY	HEMERYTHRIN	4.36e+00
15	52	65.8	196	1 K1TH_ASF7	THYMIDINE KINASE (EC 2	4.36e+00
16	52	65.8	1871	1 SEX_HUMAN	TRANSMEMBRANE PROTEIN	4.36e+00
17	51	64.6	136	1 GLB2_CALSO	GLOBIN II (HB I1)	6.79e+00
18	51	64.6	172	1 YAW7_SCHPO	VERY HYPOTHETICAL 19.6	6.79e+00
19	51	64.6	214	1 TIM3_XENLA	METALLOPROTEINASE INHI	6.79e+00
20	51	64.6	287	1 MAUN_METEX	METHYLAMINE UTILIZATIO	6.79e+00
21	51	64.6	313	1 ENVI_MOUSE	RETROVIRUS-RELATED ENV	6.79e+00
22	51	64.6	356	1 ENV_FRSEB	ENV POLYPROTEIN PRECUR	6.79e+00
23	51	64.6	408	1 ENV_FRSEFV	ENV POLYPROTEIN PRECUR	6.79e+00

24	51	64.6	409	1 ENV1_FRSEV	ENV POLYPROTEIN PRECUR	6.79e+00
25	51	64.6	409	1 ENV2_FRSEV	ENV POLYPROTEIN PRECUR	6.79e+00
26	51	64.6	567	1 PRO1_LEIEN	PROBABLE TRANSPORT PRO	6.79e+00
27	51	64.6	636	1 ENV_MCFE	ENV POLYPROTEIN PRECUR	6.79e+00
28	51	64.6	640	1 ENV_MCFE3	ENV POLYPROTEIN PRECUR	6.79e+00
29	51	64.6	724	1 NOSR_PSEST	REGULATORY PROTEIN NOS	6.79e+00
30	51	64.6	1004	1 MV10_MOUSE	PROTEIN MOV-10	6.79e+00
31	51	64.6	1057	1 TLD_DROME	DORSAL-VENTRAL PATTERN	6.79e+00
32	51	64.6	2167	1 YCS2_YEAST	HYPOTHETICAL 251.0 KD	6.79e+00
33	50	63.3	118	1 PA23_PSEAU	PHOSPHOLIPASE A2 ISOZY	1.05e+01
34	50	63.3	118	1 PA2C_PSEAU	PHOSPHOLIPASE A2 ISOZY	1.05e+01
35	50	63.3	119	1 PA22_BITNA	PHOSPHOLIPASE A2 ISOZY	1.05e+01
36	50	63.3	121	1 PA2C_BITCA	PHOSPHOLIPASE A2 (EC 3	1.05e+01
37	50	63.3	121	1 PA2C_DABRR	PHOSPHOLIPASE A2 RVV-V	1.05e+01
38	50	63.3	121	1 PA21_ERIMA	PHOSPHOLIPASE A2 ISOZY	1.05e+01
39	50	63.3	122	1 PA22_VIPAZ	PHOSPHOLIPASE A2, B CH	1.05e+01
40	50	63.3	122	1 PA2_CROAD	PHOSPHOLIPASE A2 ALPHA	1.05e+01
41	50	63.3	123	1 PA21_AKPEI	PHOSPHOLIPASE A2 (EC 3	1.05e+01
42	50	63.3	138	1 PA21_TRIFL	PHOSPHOLIPASE A2 ISOZY	1.05e+01
43	50	63.3	138	1 PA2C_CRODU	PHOSPHOLIPASE A2 CB2 P	1.05e+01
44	50	63.3	138	1 PA24_DABRU	PHOSPHOLIPASE A2 RV-4	1.05e+01
45	50	63.3	511	1 AMYS_HUMAN	ALPHA-AMYLASE SALIVARY	1.05e+01

ALIGNMENTS

RESULT	ID	NU5M_TRYBB	STANDARD;	PRT;	590 AA.
AC	P04540;				
DT	13-AUG-1987 (REL. 05, CREATED)				
DT	13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).				
GN	ND5.				
OS	TRYPANOSOMA BRUCEI BRUCEI.				
OC	MITOCHONDRION.				
OC	EUKARYOTA; EULENZOZA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE; 85037915.				
RA	HENSGENS L.A.M., BRAKENHOFF J., DE VRIES B.F., SLOOF P., TROMP M.C., VAN BOOM J.H., BENNE R.				
RA	"The sequence of the gene for cytochrome c oxidase subunit I, a frameshift containing gene for cytochrome c oxidase subunit II and seven unassigned reading frames in Trypanosoma brucei mitochondrion maxi-circle DNA."				
RL	maxi-circle DNA."				
RL	NUCLEIC ACIDS RES. 12:7327-7344(1984).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-164;				
RC	MEDLINE; 87201680.				
RA	JASMER D.P., FEAGIN J.E., PAYNE M., STUART K.;				
RA	"Variation of G-rich mitochondrial transcripts among stocks of Trypanosoma brucei."				
RL	MOL. BIOCHEM. PARASITOL. 22:259-272(1987).				
CC	-!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; X01094; ; NOT_ANNOTATED_CDS.				
DR	EMBL; M14820; G343340; ;				
DR	PIR; A04519; QOQTC5.				
DR	PFAM; PF00361; oxidored_g1; 1.				
KW	OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.				
SEQUENCE	590 AA; 71494 MW; 16042DEB CRC32;				
Query Match	70.9%; Score 56; DB 1; Length 590;				

Best Local Similarity 66.78; Pred. No. 6.96e-01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 136 AYCWE 141
QY 477 PYCWD 482

RESULT 2
ID CYGX-RAT STANDARD; PRT: 1110 AA.
AC P51839;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE OLFACTORY GUANYLYL CYCLASE GC-D PRECURSOR (EC 4.6.1.2) (GUANYLATE
CYCLASE, OLFACTORY).
GN GUCY2D.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=OLFACTORY NEURON;
RX MEDLINE; 95241343.
RA FUELLE H.-J., VASSAR R., FOSTER D.C., YANG R.-B., AXEL R.,
RA GARBERS D.L.;
RT "A receptor guanylyl cyclase expressed specifically in olfactory
sensory neurons."
RL PROC. NATL. ACAD. SCI. U.S.A. 92:3571-3575(1995).
CC -!- FUNCTION: MAY FUNCTION DIRECTLY IN ODOR RECOGNITION OR IN
MODULATING THE SENSITIVITY OF A SUBPOPULATION OF SENSORY NEURONS
TO SPECIFIC ODORS.
CC -!- CATALYTIC ACTIVITY: GTP = 3',5'-CYCLIC GMP + PYROPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN A SUBPOPULATION OF
OLFACTORY SENSORY NEURONS.
CC -!- SIMILARITY: TO OTHER GUANYLATE CYCLASES. BELONGS TO THE SENSORY
GUANYLATE CYCLASES SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L37203; G806491; -
CC PROSITE: PS00452; GUANYLATE_CYCLASES; 1.
CC DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC DR PFAM; PF00069; pkinase; 1.
CC DR PFAM; PF00211; guanylate_cyc; 1.
CC DR PFAM; PF01094; ANF_receptor; 1.
CC DR HSP; Q02846; IAWL.
CC KW LYASE; CGMP SYNTHESIS; SIGNAL; TRANSMEMBRANE; MULTIGENE FAMILY;
CC OLFACTION; GLYCOPROTEIN.
CC SIGNAL 1 66
FT CHAIN 67 1110 POTENTIAL.
FT DOMAIN 67 475 OLFACTORY GUANYLYL CYCLASE GC-D.
FT TRANSMEM 476 500 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 501 1110 POTENTIAL.
FT DOMAIN 534 819 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 822 1073 PROTEIN KINASE-LIKE.
FT DISULFID 121 149 CATALYTIC.
FT DISULFID 462 462 BY SIMILARITY.
FT DISULFID 470 470 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 304 304 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 374 374 POTENTIAL.
FT SEQUENCE 1110 AA; 121975 MW; 309B15A2 CRC32;

Query Match 70.98; Score 56; DB 1; Length 1110;
Best Local Similarity 75.08; Pred. No. 6.96e-01;
Matches 6; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 755 RDPYCSW 762
QY 475 RDP-YCGW 481

RESULT 3
ID PEX1_PICPA STANDARD; PRT: 1157 AA.
AC P46463;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 35, LAST ANNOTATION UPDATE)
DE PEROXISOME BIOSYNTHESIS PROTEIN PASI (PEROXIN-1).
GN PEX1 OR PASI.
OS PICHIA PASTORIS (YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; PICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=21-1;
RX MEDLINE; 95050987.
RA HEYMAN J.A., MONOSOV E., SUBRAMANI S.;
RT "Role of the PASI gene of Pichia pastoris in peroxisome biogenesis.";
RL J. CELL BIOL. 127:1259-1273(1994).
CC -!- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. MAY PLAY A DIRECT
OR INDIRECT ROLE IN DELIVERING MEMBRANE MATERIAL TO DEVELOPING
PEROXISOMES. IT MAY ALSO BE INVOLVED IN INTRACELLULAR MEMBRANE
MOVEMENT.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z36987; G537420; -
CC DR PROSITE; PS00674; AAA; 1.
CC DR PFAM; PF00004; AAA; 2.
CC KW PEROXISOME; ATP-BINDING; REPEAT.
FT NP_BIND 523 530 ATP (POTENTIAL).
FT NP_BIND 840 847 ATP (POTENTIAL).
FT SEQUENCE 1157 AA; 126984 MW; 3FBC3905 CRC32;

Query Match 70.98; Score 56; DB 1; Length 1157;
Best Local Similarity 71.48; Pred. No. 6.96e-01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 54 RIAYCGW 60
QY 475 RDPYCSW 481

RESULT 4
ID POLG_TUMVQ STANDARD; PRT: 3163 AA.
AC Q02597;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN (CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KD PROTEIN
DE 1 (K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2);
DE GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE (EC 3.4.22.-) (49 KD PROTEINASE) (49 KD-PRO); NUCLEAR INCLUSION
DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE COAT PROTEIN (CP1).
OS TURNIP MOSAIC VIRUS (STRAIN QUEBEC) (TUMV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE: 93057350.
RA NICOLAS O., LALIBERTE J.F.;
RT "The complete nucleotide sequence of turnip mosaic potyvirus RNA.";
RL J. GEN. VIROL. 73:2785-2793(1992).
RN (2)
RX SEQUENCE OF 1534-3163 FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE: 91073142.
RA TREMBLAY M.F., NICOLAS O., SINHA R., LAZURE C., LALIBERTE J.F.;
RT "Sequence of the 3'-terminal region of turnip mosaic virus RNA and
the capsid protein gene.";
RL J. GEN. VIROL. 71:2769-2772(1990).
RN (1)
CC -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
MAY BE INVOLVED IN REPLICATION.
CC -!- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC -!- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC -!- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
INDIVIDUAL PROTEINS.
CC -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC -!- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: D10927; G222661; -.
DR EMBL: D10601; G222659; -.
DR PIR: JQ1168; JQ1168.
DR PIR: JQ1895; JQ1895.
DR PFAM: PF00271; helicase_C; 1.
DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
DR PFAM: PF00767; Poty_coat; 1.
DR PFAM: PF00851; Peptidase_C6; 1.
DR PFAM: PF00863; Peptidase_C4; 1.
DR HYDROLASE; TRANSFERASE; THIOL PROTEASE; RNA-DIRECTED RNA POLYMERASE;
KW COAT PROTEIN; POLYPROTEIN; COVALENT PROTEIN-RNA LINKAGE; HELICASE;
KW ATP-BINDING. 1 362 N-TERMINAL PROTEIN.
FT CHAIN 363 820 HELPER COMPONENT PROTEINASE.
FT CHAIN 821 1175 PROTEIN P3.
FT CHAIN 1176 1227 6 KD PROTEIN 1.
FT CHAIN 1228 1870 CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN 1871 1923 6 KD PROTEIN 2.
FT CHAIN 1924 2115 GENOME-LINKED PROTEIN.
FT CHAIN 2116 2358 NUCLEAR INCLUSION PROTEIN A.
FT CHAIN 2359 2875 NUCLEAR INCLUSION PROTEIN B.
FT CHAIN 2876 3163 COAT PROTEIN.
FT BINDING 1986 1986 COVALENT LINKAGE OF VIRAL RNA (BY
SIMILARITY).
FT NP_BIND 1313 1320 ATP (POTENTIAL).
FT CONFLICT 2862 2862 E -> G (IN REF. 2).
SQ SEQUENCE 3163 AA; 357817 MW; 1FA8D8FA CRC32;

Query Match 70.9%; Score 56; DB 1; Length 3163;
Best Local Similarity 100.0%; Pred. No. 6.96e-01;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2586 YCGWD 2590
| | | | |
Qy 478 YCGWD 482

RESULT 5
ID POLG_TUUVJ STANDARD; PRT; 3164 AA.
AC P89509;
DT 15-DEC-1998 (REL. 37, CREATED)

DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KD PROTEIN
1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2);
GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
(EC 3.4.22.-) (49 KD PROTEINASE) (49 KD-PRO); NUCLEAR INCLUSION
PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
COAT PROTEIN (CP)].
OS TURNIP MOSAIC VIRUS (STRAIN JAPANESE) (TUMV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
RN POTYVIRUS.
CC (1)
CC SEQUENCE FROM N.A.
RX MEDLINE: 9079098.
RA OHSHIMA K., TANAKA M., SAKO N.;
RT "The complete nucleotide sequence of turnip mosaic virus RNA Japanese
strain.";
RL ARCH. VIROL. 141:1991-1997(1996).
CC -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
MAY BE INVOLVED IN REPLICATION.
CC -!- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC -!- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC -!- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
INDIVIDUAL PROTEINS.
CC -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC -!- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: D83184; G1854440; -.
DR PFAM: PF00271; helicase_C; 1.
DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
DR PFAM: PF00767; Poty_coat; 1.
DR PFAM: PF00851; Peptidase_C6; 1.
DR PFAM: PF00863; Peptidase_C4; 1.
DR HYDROLASE; TRANSFERASE; THIOL PROTEASE; RNA-DIRECTED RNA POLYMERASE;
KW COAT PROTEIN; POLYPROTEIN; COVALENT PROTEIN-RNA LINKAGE; HELICASE;
KW ATP-BINDING. 1 362 N-TERMINAL PROTEIN.
FT CHAIN 363 820 HELPER COMPONENT PROTEINASE.
FT CHAIN 821 1175 PROTEIN P3.
FT CHAIN 1176 1227 6 KD PROTEIN 1.
FT CHAIN 1228 1871 CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN 1872 1924 6 KD PROTEIN 2.
FT CHAIN 1925 2116 GENOME-LINKED PROTEIN.
FT CHAIN 2117 2359 NUCLEAR INCLUSION PROTEIN A.
FT CHAIN 2360 2876 NUCLEAR INCLUSION PROTEIN B.
FT CHAIN 2877 3164 COAT PROTEIN.
FT BINDING 1987 1987 COVALENT LINKAGE OF VIRAL RNA (BY
SIMILARITY).
FT NP_BIND 1313 1320 ATP (POTENTIAL).
SQ SEQUENCE 3164 AA; 357734 MW; 6406E37A CRC32;

Query Match 70.9%; Score 56; DB 1; Length 3164;
Best Local Similarity 100.0%; Pred. No. 6.96e-01;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2587 YCGWD 2591
| | | | |
Qy 478 YCGWD 482

RESULT 6

```

ID AC RRPL_SV5WR STANDARD: PRT: 2255 AA.
DT Q03396;
DT 01-FEB-1994 (REL. 28, LAST CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE (L PROTEIN).
DE L.
OS SIMIAN VIRUS 5 (STRAIN 21004-WR) (SV5).
OC VIRUSES: SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92341049.
RA HIGUCHI Y., MIYAHARA Y., KAWANO M., TSURUDOME M., MATSUMURA H.,
RA KUSAGAWA S., KONADA H., NISHIO M., ITO Y.;
RT "Sequence analysis of the large (L) protein of simian virus 5.";
RL J. GEN. VIROL. 73:1005-1010(1992).
CC -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs, RNA EDITING OF THE P
CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D13868; G222628; -.
DR PIR: J01532; J01532.
DR PFAM: PF00946; Paramyx_RNA_pol; 1.
DR TRANSFERASE; RNA-DIRECTED RNA POLYMERASE.
KW SEQUENCE 2255 AA; 255938 MW; 3A882A5F CRC32;
SQ SEQUENCE 2255 AA; 255938 MW; 3A882A5F CRC32;

Query Match 69.6%; Score 55; DB 1; Length 2255;
Best Local Similarity 80.0%; Pred. No. 1.11e+00;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1490 YCGWE 1494
QY 478 YCGWD 482
|||||

RESULT 7 STANDARD: PRT: 2255 AA.
ID AC RRPL_SV5
DT Q88434;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE (L PROTEIN).
DE L.
OS SIMIAN VIRUS 5 (STRAIN W3) (SV5).
OC VIRUSES: SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92327825.
RA PARKS G.D., WARD C.D., LAMB R.A.;
RT "Molecular cloning of the NP and L genes of simian virus 5:
RT Identification of highly conserved domains in paramyxovirus NP and L
RT proteins.";
RL VIRUS RES. 22:259-279(1992).
CC -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs, RNA EDITING OF THE P
CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D13868; G222628; -.
DR PIR: J01532; J01532.
DR PFAM: PF00946; Paramyx_RNA_pol; 1.
DR TRANSFERASE; RNA-DIRECTED RNA POLYMERASE.
KW SEQUENCE 2255 AA; 255938 MW; 3A882A5F CRC32;
SQ SEQUENCE 2255 AA; 255938 MW; 3A882A5F CRC32;

Query Match 69.6%; Score 55; DB 1; Length 2255;
Best Local Similarity 80.0%; Pred. No. 1.11e+00;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1490 YCGWE 1494
QY 478 YCGWD 482
|||||

RESULT 8 STANDARD: PRT: 286 AA.
ID AC AAC8_STRFR
DT P29809;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE AMINOGLYCOSIDE 3'-N-ACETYLTRANSFERASE VIII (EC 2.3.1.81) (ACC(3)-VIII)
DE (AMINOGLYCOSIDE 3'-N-ACETYLTRANSFERASE TYPE VIII).
DE GN AAC8.
OS STREPTOMYCES FRADIAE.
OC BACTERIA: FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE.
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91285425.
RA SALAUZE D., PEREZ-GONZALEZ J.A., PIEPERSBERG W., DAVIES J.;
RT "Characterisation of aminoglycoside acetyltransferase-encoding genes
RT of neomycin-producing Micromonospora chalicea and Streptomyces
RT fradiae.";
RL GENE 101:143-148(1991).
CC -!- FUNCTION: RESISTANCE TO NEOMYCIN.
CC -!- CATALYTIC ACTIVITY: ACETYL-COA + A 2-DEOXYSTREPTAMINE ANTIBIOTIC
CC = COA + N3'-ACETYL-2-DEOXYSTREPTAMINE ANTIBIOTIC.
CC -!- SIMILARITY: TO OTHER AAC(3) PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M55426; G295164; -.
DR PIR: JE0417; JE0417.
DR TRANSFERASE; TRANSFERASE; ACYLTRANSFERASE.
KW ANTIBIOTIC RESISTANCE;
SQ SEQUENCE 286 AA; 30455 MW; 44C39995 CRC32;

Query Match 68.4%; Score 54; DB 1; Length 286;
Best Local Similarity 66.7%; Pred. No. 1.76e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 69 AYCGWN 74
QY 477 PYCGWD 482
|||||

RESULT 9 STANDARD: PRT: 417 AA.
ID AC AGP_PROBE
DT Q52309;

```



```

DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GLUCOSE-1-PHOSPHATASE PRECURSOR (EC 3.1.3.10) (GIPASE).
GN AGP.
OS PROVIDENCIA RETTGERTI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC PROVIDENCIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PV7;
RA RICCIO M.L., CHIESURIN A., LOMBARDI G., SATTA G.;
RL SUBMITTED (SEP-1992) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: D-GLUCOSE 1-PHOSPHATE + H(2)O -> D-GLUCOSE +
CC ORTHOPHOSPHATE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68201; G45772; -
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
DR PFAM; PF00328; acid_phosphat; 1.
KW HYDROLASE; PERIPLASMIC; SIGNAL.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 417 GLUCOSE-1-PHOSPHATASE.
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT SITE 314 314 PROTON DONOR (BY SIMILARITY).
FT SEQUENCE 417 AA; 46940 MW; 4C5778A5 CRC32;
Query Match 68.4%; Score 54; DB 1; Length 417;
Best Local Similarity 37.5%; Pred. No. 1.76e+00;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 398 KNGCSWE 405
Qy 475 RDPYCGWD 482
RESULT 10
ID UL04_HSVB STANDARD; PRT; 225 AA.
AC P28943;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE GENE 58 PROTEIN.
GN 58
OS EQUINE HERPESVIRUS TYPE 1 (STRAIN AB4P) (EHV-1).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 92295566.
RA TELFORD E.A.R., WATSON M.S., MCBRIDE K., DAVISON A.J.;
RL "The DNA sequence of equine herpesvirus-1.";
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL4,
CC HSV-2 UL4, EHV-1 58, AND VZV 56.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86664; G330849; -
DR PIR; D36801; WZBEF1.
SQ SEQUENCE 225 AA; 24287 MW; 613DF9DB CRC32;
Query Match 67.1%; Score 53; DB 1; Length 225;
Best Local Similarity 80.0%; Pred. No. 2.78e+00;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 104 PFCGW 108
Qy 477 PYCGW 481
RESULT 11
ID RPO_PVSP STANDARD; PRT; 353 AA.
AC P22657;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RNA REPLICATION PROTEIN (223 KD PROTEIN) (ORF 1) [CONTAINS: RNA-
DE DIRECTED RNA POLYMERASE (EC 2.7.7.48); PROBABLE HELICASE] (FRAGMENT).
OS POTATO VIRUS S (STRAIN PERUVIAN).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CARLAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 89279283.
RA MACKENZIE D.J., TREMAINE J.H., STACE-SMITH R.;
RT "Organization and intervirial homologies of the 3'-terminal portion of
RT potato virus S RNA.";
RL J. GEN. VIROL. 70:1053-1063(1989).
CC -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
CC POSSIBLY FUNCTIONS AS A ATP-BINDING HELICASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00461; G222439; -
DR PIR; JAO123; JAO123.
KW ATP-BINDING; HELICASE; RNA REPLICATION; RNA-DIRECTED RNA POLYMERASE;
KW TRANSFERASE.
FT NON_TER 1 1
FT SEQUENCE 353 AA; 41063 MW; A3C18F6C CRC32;
Query Match 67.1%; Score 53; DB 1; Length 353;
Best Local Similarity 57.1%; Pred. No. 2.78e+00;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 262 RPSFCGW 268
Qy 475 RDPYCGW 481
RESULT 12
ID GLTD_SCOLI STANDARD; PRT; 471 AA.
AC P09832;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUTAMATE SYNTHASE [NADPH] SMALL CHAIN (EC 1.4.1.13) (GLUTAMATE
DE SYNTHASE BETA SUBUNIT) (NADPH-GOGAT) (GLTS BETA CHAIN).
GN GLTD OR ASPB.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
```

```

FT CONFLICT      375      399      GRRRIIVAGSEHIVPADAVIMAFG -> ASPRGDCRCRFMT
SQ SEQUENCE      471 AA; 51884 MW; FE00B193 CRC32;
                                         YRTGCGDGRGVW (IN REF. 1).

Query Match          67.1% Score 53; DB 1; Length 471;
Best Local Similarity 56.7%; Pred.No. 2.78e+00;
Matches              4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db       51 NPYCEW 56
QY       476 DPYCGW 481
        :||| |
RESULT 13
ID ID YC8L CAEEL STANDARD; PRT; 1291 AA.
AC AC Q19981;
DT DT 01-NOV-1997 (REL. 35, CREATED)
DT DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DE HYPOTHETICAL 143.1 KD PROTEIN F33C8.1 IN CHROMOSOME X PRECURSOR.
GN GN F33C8.1.
OS OS CAENORHABDITIS ELEGANS.
OC OC UKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN RN [1]
RP RP SEQUENCE FROM N.A.
RC RC STRAIN-BRISTOL N2;
RA RA PERCY C.;
RL RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC CC -!- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
-----
CC CC EMBL; Z69790; EI346241; -.
DR DR WORMPEP; F33C8.1; CE05796.
DR DR PROSITE; PS00022; EGF_1; 2.
DR DR PROSITE; PS01186; EGF_2; 2.
DR DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR DR PROSITE; PS01180; CUB; 1.
DR DR PFAM; PF00008; EGF; 2.
DR DR PFAM; PF00053; laminin_EGF; 2.
DR DR PFAM; PF00431; CUB; 1.
DR DR PFAM; PF01344; Kelch; 2.
DR DR HSP; P02458; IKLO.
KW KW HYPOTHETICAL PROTEIN; LAMININ EGF-LIKE DOMAIN; EGF-LIKE DOMAIN; REPEAT; SIGNAL.
FT FT SIGNAL 1 ? POTENTIAL.
FT FT CHAIN ? 1291 HYPOTHETICAL 143.1 KD PROTEIN.
FT FT DOMAIN 62 92 EGF-LIKE 1.
FT FT DOMAIN 94 200 CUB.
FT FT DOMAIN 201 232 EGF-LIKE 2.
FT FT DOMAIN 232 270 EGF-LIKE 3.
FT FT DOMAIN 965 1019 LAMININ EGF-LIKE 1.
FT FT DOMAIN 1020 1067 LAMININ EGF-LIKE 2.
FT FT DISULFID 66 75 BY SIMILARITY.
FT FT DISULFID 70 80 BY SIMILARITY.
FT FT DISULFID 82 91 BY SIMILARITY.
FT FT DISULFID 205 215 BY SIMILARITY.
FT FT DISULFID 209 220 BY SIMILARITY.
FT FT DISULFID 222 231 BY SIMILARITY.
FT FT DISULFID 236 252 BY SIMILARITY.
FT FT DISULFID 247 257 BY SIMILARITY.
FT FT DISULFID 259 269 BY SIMILARITY.
FT FT DISULFID 965 973 BY SIMILARITY.
FT FT DISULFID 967 988 BY SIMILARITY.

```

FT DISULFID 991 1000 BY SIMILARITY.
FT DISULFID 1003 1017 BY SIMILARITY.
FT DISULFID 1020 1029 BY SIMILARITY.
FT DISULFID 1022 1036 BY SIMILARITY.
FT DISULFID 1038 1048 BY SIMILARITY.
FT DISULFID 1051 1065 BY SIMILARITY.
FT CARBOHYD 103 103 POTENTIAL.
FT CARBOHYD 197 197 POTENTIAL.
FT CARBOHYD 208 208 POTENTIAL.
FT CARBOHYD 324 324 POTENTIAL.
FT CARBOHYD 395 395 POTENTIAL.
FT CARBOHYD 447 447 POTENTIAL.
FT CARBOHYD 481 481 POTENTIAL.
FT CARBOHYD 529 529 POTENTIAL.
FT CARBOHYD 555 555 POTENTIAL.
FT CARBOHYD 840 840 POTENTIAL.
FT CARBOHYD 853 853 POTENTIAL.
FT CARBOHYD 954 954 POTENTIAL.
FT CARBOHYD 993 993 POTENTIAL.
FT CARBOHYD 1086 1086 POTENTIAL.
FT CARBOHYD 1122 1122 POTENTIAL.
FT CARBOHYD 1167 1167 POTENTIAL.
SQ SEQUENCE 1291 AA; 143052 MW; 687FD8C0 CRC32;

Query Match 67.1%; Score 53; DB 1; Length 1291;
Best Local Similarity 85.7%; Pred. No. 2.78e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 921 RDPGCGW 927
QY 475 RDPYCGW 481
||| |||
RESULT 14
ID HEMT_THEDY STANDARD; PRT; 113 AA.
AC P02245;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HEMERYTHRIN.
OS THEMISTE DYSCRITUM.
OC EUKARYOTA; METAZOA; SIPUNCULA; GOLFGINGIIDAE; THEMISTE.
RN [1]
RP SEQUENCE.
RX MEDLINE; 78218273.
RA LOEHR J.S., LAMMERS P.J., BRIMHALL B., HERMOSON M.A.;
RT "Amino acid sequence of hemerythrin from Themiste dyscritum.";
RL J. BIOL. CHEM. 253:5746-5731(1978).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND IRON LIGANDS.
RX MEDLINE; 78235906.
RA STENKAMP R.E., SIEKER L.C., JENSEN L.H., MCQUEEN J.E. JR.;
RT "Structure of methemerythrin at 2.8-A resolution: computer graphics
fit of an averaged electron density map.";
RL BIOCHEMISTRY 17:2499-2504(1978).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS), AND IRON LIGANDS.
RA STENKAMP R.E., SIEKER L.C., JENSEN L.H., SANDERS-LOEHR J.;
RT "Structure of the binuclear iron complex in metazidohaemerythrin from
Themiste dyscritum at 2.2-A resolution.";
RL NATURE 291:263-264(1981).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.66 ANGSTROMS).
RX MEDLINE; 91332904.
RA HOLMES M.A., STENKAMP R.E.;
RT "Structures of met and azidomet hemerythrin at 1.66-A resolution.";
RL J. MOL. BIOL. 220:723-737(1991)
CC -1- FUNCTION: HEMERYTHRIN IS A RESPIRATORY PROTEIN IN BLOOD CELLS OF
CERTAIN MARINE WORMS. THE OXYGEN-BINDING SITE IN EACH CHAIN
CONTAINS TWO IRON ATOMS.
CC -1- SUBUNIT: HOMOOCTAMER.
CC -1- SIMILARITY: BELONGS TO THE HEMERYTHRIN FAMILY.
DR PIR; A02572; HRTHBD.

DR PDB; 1HMD; 15-JUL-92.
DR PDB; 1HMO; 15-JUL-92.
DR PDB; 2HMZ; 15-JAN-92.
DR PDB; 2HMZ; 15-JAN-92.
DR PROSITE; PS00550; HEMERYTHRINS; 1.
KW OXYGEN TRANSPORT; METAL-BINDING; IRON; 3D-STRUCTURE.
FT METAL 25 25 IRON 1.
FT METAL 54 54 IRON 1.
FT METAL 58 58 IRON 1 AND 2.
FT METAL 73 73 IRON 2.
FT METAL 77 77 IRON 2.
FT METAL 101 101 IRON 2.
FT METAL 106 106 IRON 1 AND 2.
FT VARIANT 64 64 A -> S.
FT HELIX 12 14
FT TURN 19 21
FT HELIX 22 37
FT HELIX 41 64
FT TURN 65 66
FT TURN 68 69
FT HELIX 70 85
FT TURN 86 86
FT HELIX 91 104
FT TURN 105 105
FT HELIX 106 109
FT TURN 110 112
SQ SEQUENCE 113 AA; 13319 MW; EBC9C9D3 CRC32;

Query Match 65.8%; Score 52; DB 1; Length 113;
Best Local Similarity 85.7%; Pred. No. 4.36e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 6 DPYC-WD 11
QY 476 DPYCGWD 482
||| |||
RESULT 15
ID KITH_ASFB7 STANDARD; PRT; 196 AA.
AC P18555;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE THYMIDINE KINASE (EC 2.7.1.21).
GN TDK OR K196R.
OS AFRICAN SWINE FEVER VIRUS (STRAIN BA71V) (AFSV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE;
OC AFRICAN SWINE FEVER-LIKE VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90357780.
RA BLASCO R., LOPEZ-OTIN C., MUNOZ M., BOCKAMP E.-O., SIMON-MATEO C.,
RA VINUELA E.;
RT "Sequence and evolutionary relationships of African swine fever virus
thymidine kinase.";
RL VIROLOGY 178:301-304(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91101261.
RA HERNANDEZ A.M., TABERES E.;
RT "Expression and characterization of the thymidine kinase gene of
African swine fever virus.";
RL J. VIROL. 65:1046-1052(1991).
RN [3]
RP COMPLETE GENOME.
RA YANEZ R.J., RODRIGUEZ J.M., NOGAL M.L., YUSTE L., ENRIQUEZ C.,
RA RODRIGUEZ J.F., VINUELA E.;
RT "Analysis of the complete nucleotide sequence of African swine fever
virus.";
RL VIROLOGY 208:249-278(1995).
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE -> ADP + THYMIDINE
CC 5'-PHOSPHATE.
CC -1- SIMILARITY: TO BOTH POXVIRUSES AND VERTEBRATE THYMIDINE KINASES.

```

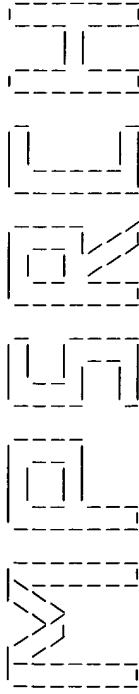
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31715; G210656; -
DR EMBL; M63119; G210658; -
DR EMBL; U18486; G780420; -
DR PIR; A37079; KIXFAS.
DR PROSITE; PS00603; TK_CELLULAR_TYPE; 1.
DR PFAM; PF00265; TK; 1.
KW TRANSFERASE; KINASE; DNA SYNTHESIS; ATP-BINDING.
FT NP_BIND 17 24 ATP (PROBABLE).
FT BINDING 45 45 ATP (PROBABLE).
SQ SEQUENCE 196 AA; 22420 MW; 340EA74C CRC32;

Query Match      65.8%; Score 52; DB 1; Length 196;
Best Local Similarity 80.0%; Pred. No. 4.36e+00;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 134 PYCSW 138
QY 477 PYCGW 481

```

Search completed: Thu Jul 8 18:57:24 1999
Job time : 6 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 19:00:48 1999; MasPar time 10.62 Seconds
32.052 Million cell updates/sec
Tabular output not generated.

Title: >US-09-041-236-2
Description: (480-495) from US09041236.pgp (26 of 45)
Perfect Score: 124
Sequence: 1 GWDQGRCSISYSSERS 16

Scoring table:
PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 20.255; Variance 62.606; scale 0.324

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	66	53.2	474	32	Variant of GABA-A rec	9.19e+00
2	66	53.2	506	32	Human GABA receptor e	9.19e+00
3	66	53.2	506	32	GABA-A receptor epsil	9.19e+00
4	62	50.0	455	29	Human kidney calcium	2.42e+01
5	62	50.0	455	29	Human placental calci	2.42e+01
6	62	50.0	455	17	Human calcium sensor	2.42e+01
7	62	50.0	455	29	Human calcium sensor	2.42e+01
8	62	50.0	455	17	Human placental calci	2.42e+01
9	62	50.0	455	29	Human parathyroid cal	2.42e+01
10	62	50.0	455	17	Human parathyroid cal	2.42e+01
11	62	50.0	455	17	Human kidney calcium	2.42e+01
12	60	48.4	339	3	Sequence of the last	3.89e+01
13	60	48.4	341	5	Sequence encoded by t	3.89e+01
14	60	48.4	524	18	DHFR/HPH fusion prote	3.89e+01
15	60	48.4	539	18	DHFR/HPH fusion prote	3.89e+01
16	60	48.4	539	18	DHFR/HPH fusion prote	3.89e+01

17	60	48.4	691	4	R24024	Sequence encoded by t	3.89e+01
18	58	45.8	3011	16	R95021	Hepatitis GB virus (H	6.23e+01
19	57	46.0	717	1	P91933	BI antigen.	7.86e+01
20	57	46.0	1023	3	R13319	Partial Human Natural	7.86e+01
21	57	46.0	1358	33	W57838	Aldehyde oxidase prot	7.86e+01
22	57	46.0	1358	33	W57837	Aldehyde oxidase prot	7.86e+01
23	55	44.4	132	15	R82987	Human transforming gr	1.25e+02
24	55	44.4	152	12	R74824	Bacteriophage T3 SAM-	1.25e+02
25	55	44.4	152	16	R88601	S-adenosylmethionine	1.25e+02
26	55	44.4	152	14	R75212	Bacteriophage T3 S-ad	1.25e+02
27	55	44.4	152	23	W11594	AdoMetase.	1.25e+02
28	55	44.4	152	3	R12520	Adometase encoded by	1.25e+02
29	55	44.4	152	25	W29147	S-adenosylmethionine	1.25e+02
30	55	44.4	152	16	R88603	Adometase (I2V) SAM-K	1.25e+02
31	55	44.4	152	17	R88611	S-adenosylmethionine-	1.25e+02
32	55	44.4	152	17	R88606	S-adenosylmethionine-	1.25e+02
33	55	44.4	380	23	W09406	Transforming growth f	1.25e+02
34	55	44.4	551	24	W27150	HMG-CoA reductase deg	1.25e+02
35	55	44.4	1274	7	R34714	Bacillus subtilis srf	1.25e+02
36	54	43.5	125	25	W32372	Mycobacterium tubercu	1.57e+02
37	54	43.5	125	37	W81675	M. tuberculosis immun	1.57e+02
38	54	43.5	125	25	W32440	Mycobacterium tubercu	1.57e+02
39	54	43.5	125	35	W64312	Mycobacterium tubercu	1.57e+02
40	54	43.5	610	19	W03422	Mouse developmental k	1.57e+02
41	54	43.5	626	19	W03423	Mouse developmental k	1.57e+02
42	54	43.5	998	15	R85092	EPH-like receptor pro	1.57e+02
43	54	43.5	998	19	W03421	Mouse developmental k	1.57e+02
44	53	42.7	375	16	R0919	Connective tissue gro	1.97e+02
45	53	42.7	404	34	W62620	Mus musculus SOCS6 pr	1.97e+02

ALIGNMENTS

RESULT 1
ID W59049 standard; Protein; 474 AA.
AC W59049;
DT 25-SEP-1998 (first entry)
DE Variant of GABA-A receptor epsilon sub-unit related protein.
KW GABA-A receptor; diagnosis; therapy; drug screening.
OS Mammalia.
PN DE19644501-A1.
PD 30-APR-1998.
PF 25-OCT-1996; 044501.
PR 25-OCT-1996; DE-044501.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI Bilke K, Gaul R, Kioschis P, Poustka A, Wilke K;
DR WPI; 98-252096/23.
DR N-PSDB; V11777.
PT Protein related to GABA-A receptor sub-unit epsilon - useful for
PT diagnosis, therapy or drug screening
PS Disclosure; Fig 2; 12pp; German.
CC The GABA-A receptor sub-unit epsilon related protein is used for the
CC diagnosis and/or therapy of related disorders. It can also be used to
CC screen for substances that modulate the activity of the protein.
SQ Sequence 474 AA;

Query Match 53.2%; Score 66; DB 32; Length 474;
Best Local Similarity 63.6%; Pred. No. 9.19e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;
Db 436 wqggrlchvy 446
Qy 481 WQDGR-CISYI 490

RESULT 2
ID W61045 standard; Protein; 506 AA.
AC W61045;
DT 12-OCT-1998 (first entry)
DE Human GABA receptor epsilon subunit.
KW GABA receptor; gamma-amino butyric acid receptor; human;
KW appetite; cognition.
OS Homo sapiens.

PN W09823742-A1.
PD 04-JUN-1998.
PF 18-NOV-1997; G03159.
PR 03-OCT-1997; GB-020995.
PR 25-NOV-1996; GB-024442.
PA (MERI) MERCK SHARP & DOHME LTD.
PI Whiting FJ;
DR WPI: 98-322722/28.
DR N-PSDB: V36319.
PT New isolated GABA receptor subunit, epsilon - used to develop
PT products for the screening and design of drugs, e.g. for modulating
PT appetite behaviours, hormonal interactions and cognition
PS Claim 15; Fig 1; 37pp; English.
CC This polypeptide comprises a novel class of subunit, classified as
CC as epsilon, within the human GABA receptor gene family. Its amino
CC acid sequence was deduced from a cDNA clone (see V36319) isolated
CC from a human brain cDNA library. A stably co-transfected eukaryotic
CC cell line capable of expressing the GABA epsilon receptor subunit,
CC at least one alpha receptor subunit and at least one beta receptor
CC subunit is claimed. A recombinant expression vector, and protein
CC and membrane preparations of the GABA receptor obtained from the
CC co-transfected cells are also claimed. The new GABA-A receptor
CC epsilon subunit and products can be used in the screening and
CC design of drugs which act upon the GABA receptor. The ligands or
CC drug candidates identified may be used as therapeutic agents, e.g.
CC for the modulation of appetite behaviours, hormonal interactions
CC and cognition. They can also be used for the study of structure,
CC pharmacology and function of the various GABA receptor subtypes.
SQ Sequence 506 AA;

Query Match 53.2%; Score 66; DB 33; Length 506;
Best Local Similarity 63.6%; Pred. No. 9.19e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 468 wqggrlcihvy 478
I:|||||:
Qy 481 WDOGR-CISY 490

RESULT 3
ID W9048 standard; Protein; 506 AA.
AC W59048;
DT 25-SEP-1998 (first entry)
DE GABA-A receptor epsilon sub-unit related protein.
KW GABA-A receptor; diagnosis; therapy; drug screening.
OS Mammalia
PN DE19644501-A1.
PD 30-APR-1998.
PF 25-OCT-1996; 044501.
PR 25-OCT-1996; DE-044501.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI Bläke K, Gaul R, Kioschis P, Poustka A, Wilke K;
DR WPI: 98-252096/23.
DR N-PSDB: V11776.
PT Protein related to GABA-A receptor sub-unit epsilon - useful for
PT diagnosis, therapy or drug screening
PS Example 1; Fig 1; 12pp; German.
CC The GABA-A receptor sub-unit epsilon related protein is used for the
CC diagnosis and/or therapy of related disorders. It can also be used to
CC screen for substances that modulate the activity of the protein.
SQ Sequence 506 AA;

Query Match 53.2%; Score 66; DB 32; Length 506;
Best Local Similarity 63.6%; Pred. No. 9.19e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 468 wqggrlcihvy 478
I:|||||:
Qy 481 WDOGR-CISY 490

RESULT 4
ID W43313 standard; Protein; 4655 AA.

AC W43313;
DT 08-JUN-1998 (first entry)
DE Human kidney calcium sensor protein.
KW Calcium sensor; human; glycoprotein 330; gp330; kidney;
KW low density lipoprotein; LDL; membranous glomerulonephritis;
KW osteoporosis; hyperthyroidism; therapy.
OS Homo sapiens.
FH Key
FT Peptide
1..25
/label= Sig_peptide
FT Domain
26..4422
/notes= "extracellular domain"
FT Domain
4423..4445
/notes= "transmembrane domain"
FT Domain
4446..4655
/notes= "C-terminal domain"
FT Misc_difference 4210
/notes= "encoded by MTC"
FT W09744050-A1.
PN 27-NOV-1997.
PD 20-MAY-1997; U08682.
PR 23-MAY-1996; US-652877.
PA (RHON) RHONE-POULENC RORER PHARM INC.
PI Akerstrom G, Crumley GR, Hjalml G, Juhlml C, Morse CC,
PI Murray EM, Rask L;
DR WPI: 98-018219/02.
DR N-PSDB: V05996;
PT New human calcium sensor protein and vector encoding inhibitor of gp
PT 330 mediated transport of amyloid beta protein - useful to treat
PT membranous glomerulonephritis and central nervous system disorders
PS Claim 7; Page 145-163; 240pp; English.
CC This sequence comprises a human kidney calcium sensor protein
CC (hCSP), deduced from the nucleotide sequence (see V05996) of a
CC clone obtained by PCR amplification. Full-length sequences for
CC human placenta (see W43312) and parathyroid (see W43314) are also
CC claimed. hCSP is considered to be the human homologue of rat
CC Heymann nephritis antigen gp330 on the basis of sequence
CC similarity. It is proposed that the calcium sensor is a
CC universal sensor for recognition of variation in extracellular
CC calcium, and plays a key role in calcium regulation via different
CC organ systems. hCSP belongs to the low density lipoprotein
CC superfamily of glycoproteins. It can be expressed in a claimed
CC host cell using a vector comprising hCSP cDNA. Also claimed are a
CC method of using hCSP to treat human membranous glomerulonephritis,
CC and the use of hCSP to screen for agonists and antagonists of hCSP.
CC Such (ant)agonists are useful in the treatment of conditions
CC associated with perturbations in parathyroid hormone levels,
CC vitamin D3 production, oestrogen levels, osteoclast or osteoblast
CC activity, calcium secretion and calcium ion homeostasis, e.g.
CC hyperthyroidism and osteoporosis.
SQ Sequence 4655 AA;

Query Match 50.0%; Score 62; DB 29; Length 4655;
Best Local Similarity 77.8%; Pred. No. 2.42e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 275 esgrclsiy 283
:|||||
Qy 482 DOGRCSY 490

RESULT 5
ID W43312 standard; Protein; 4655 AA.
AC W43312;
DT 08-JUN-1998 (first entry)
DE Human placental calcium sensor protein.
KW Calcium sensor; human; glycoprotein 330; gp330; placenta;
KW low density lipoprotein; LDL; membranous glomerulonephritis;
KW osteoporosis; hyperthyroidism; therapy.
OS Homo sapiens.
FH Key
FT Peptide
1..25
/label= Sig_peptide
FT

```

FT Domain 26..4422
FT /note= "extracellular domain"
FT Domain 4423..4445
FT /note= "transmembrane domain"
FT Domain 4446..4655
FT /note= "C-terminal domain"
FT Misc_difference 213
FT /note= "encoded by GAY"
FT Misc_difference 675
FT /note= "encoded by GTY"
FT Misc_difference 1272
FT /note= "encoded by CCH"
FT Misc_difference 1287
FT /note= "encoded by SCA"
FT Misc_difference 4094
FT /note= "encoded by RAG"
FT Misc_difference 4210
FT /note= "encoded by MTC"
FT Misc_difference 4371
FT /note= "encoded by ATY"
PN WO9744050-A1.
PD 27-NOV-1997.
PF 20-MAY-1997; U08682.
PR 23-MAY-1996; US-652877.
PA (RHON ) RHONE-POULENC RORER PHARM INC.
PI Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC,
PI Murray EM, Rask L;
DR WPI: 98-018219/02.
DR N-PSDB; V05995.
PT New human calcium sensor protein and vector encoding inhibitor of gp
PT 330 mediated transport of amyloid beta protein - useful to treat
PT membranous glomerulonephritis and central nervous system disorders
PS Claim 7; Page 112-130; 240pp; English.
CC This sequence comprises a human placental calcium sensor protein
CC (hCSP), deduced from the nucleotide sequence (see V05995) of a
CC clone obtained by PCR amplification. Full-length sequences for
CC human kidney (see W43313) and parathyroid (see W43314) are also
CC claimed. hCSP is considered to be the human homologue of rat
CC Heymann nephritis antigen gp330 on the basis of sequence
CC similarity. It is proposed that the calcium sensor is a
CC universal sensor for recognition of variation in extracellular
CC calcium, and plays a key role in calcium regulation via different
CC organ systems. hCSP belongs to the low density lipoprotein
CC superfamily of glycoproteins. It can be expressed in a claimed
CC host cell using a vector comprising hCSP cDNA. Also claimed are a
CC method of using hCSP to treat human membranous glomerulonephritis,
CC and the use of hCSP to screen for agonists and antagonists of hCSP.
CC Such (ant)agonists are useful in the treatment of conditions
CC associated with perturbations in parathyroid hormone levels,
CC vitamin D3 production, oestrogen levels, osteoclast or osteoblast
CC activity, calcium secretion and calcium ion homeostasis, e.g.
CC hyperthyroidism and osteoporosis.
SQ Sequence 4655 AA;

Query Match 50.0%; Score 62; DB 29; Length 4655;
Best Local Similarity 77.8%; Pred. No. 2.42e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 275 esgrcisiy 283
Qy 482 DOGRCSIIY 490

RESULT 6
ID R97208 standard; Protein; 4655 AA.
AC R97208;
DE 31-JUL-1996 (first entry)
KW Human calcium sensor protein.
KW Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;
KW membranous glomerulonephritis; therapy; ulcer; wound; kidney.
OS Homo sapiens.
PN WO9615801-A1.
PD 30-MAY-1996.

```

```

PF 22-NOV-1995; U15203.
PR 23-NOV-1994; US-344836.
PR 07-JUN-1995; US-487314.
PA (RHON ) RHONE-POULENC RORER PHARM INC.
PI Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC;
PI Murray EM, Rask L;
DR WPI: 96-268334/27.
DR N-PSDB; T28773.
PT Isolated human calcium sensor protein - useful develop prods. for
PT diagnosis and treatment of conditions, such as membranous
PT glomerulonephritis
PS Claim 7; Page 78-96; 236pp; English.
CC A 500 kDa human calcium sensor protein (CSP) (R97208) was identified
CC as the product of cDNA clone (T28773) obt'd. from human kidney and
CC placental cDNA libraries. Full-length sequences of placental, kidney
CC and parathyroid CSP (see also R97209-11) were also obt'd. They show
CC homology to LDL receptor proteins. CSP can be isolated from
CC parathyroid, placental or kidney tubule cells or obt'd. by expression of
CC encoding sequences in transformed host cells. It is useful for
CC identifying potential (ant)agonists and in the diagnosis and treatment
CC of e.g. membranous glomerulonephritis, and also in the treatment of
CC wounds and ulcers.
SQ Sequence 4655 AA;

Query Match 50.0%; Score 62; DB 17; Length 4655;
Best Local Similarity 77.8%; Pred. No. 2.42e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 275 esgrcisiy 283
Qy 482 DOGRCSIIY 490

RESULT 7
ID W43311 standard; Protein; 4655 AA.
AC W43311;
DE 08-JUN-1998 (first entry)
DE Human calcium sensor protein.
KW Calcium sensor; human; glycoprotein 330; gp330; placenta; kidney;
KW low density lipoprotein; LDL; membranous glomerulonephritis;
KW osteoporosis; hyperthyroidism; therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..25
FT /label= Sig_peptide
FT Domain 26..4422
FT /note= "extracellular domain"
FT Domain 4423..4445
FT /note= "transmembrane domain"
FT Domain 4446..4655
FT /note= "C-terminal domain"
PN WO9744050-A1.
PD 27-NOV-1997.
PF 20-MAY-1997; U08682.
PR 23-MAY-1996; US-652877.
PA (RHON ) RHONE-POULENC RORER PHARM INC.
PI Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC,
PI Murray EM, Rask L;
DR WPI: 98-018219/02.
DR N-PSDB; V05994.
PT New human calcium sensor protein and vector encoding inhibitor of gp
PT 330 mediated transport of amyloid beta protein - useful to treat
PT membranous glomerulonephritis and central nervous system disorders
PS Claim 7; Page 80-98; 240pp; English.
CC This sequence comprises a human calcium sensor protein (hCSP),
CC deduced from the nucleotide sequence (see V05994) of clones
CC isolated from a human placental and kidney libraries. Full-length
CC sequences (see W43312-14) for human kidney, placenta and
CC parathyroid are also claimed. hCSP is considered to be the human
CC homologue of rat Heymann nephritis antigen gp330 on the basis of
CC sequence similarity. It is proposed that the calcium sensor is a
CC universal sensor for recognition of variation in extracellular
CC calcium, and plays a key role in calcium regulation via different

```

CC organ systems. hcsp belongs to the low density lipoprotein
 CC superfamily of glycoproteins. It can be expressed in a claimed
 CC host cell using a vector comprising hcsp cDNA. Also claimed are a
 CC method of using hcsp to treat human membranous glomerulonephritis,
 CC and the use of hcsp to screen for agonists and antagonists of hcsp.
 CC Such (ant)agonists are useful in the treatment of conditions
 CC associated with perturbations in parathyroid hormone levels,
 CC vitamin D3 production, oestrogen levels, osteoclast or osteoblast
 CC activity, calcium secretion and calcium ion homeostasis, e.g.
 CC hyperthyroidism and osteoporosis.
 SQ Sequence 4655 AA;

Query Match 50.0%; Score 62; DB 29; Length 4655;
 Best Local Similarity 77.8%; Pred. No. 2.42e+01;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 275 esgcrisiiy 283
 : |||||

Qy 482 DQGRCSIIY 490

RESULT 8
 ID R97209 standard; Protein; 4655 AA.
 AC R97209;
 DT 31-JUL-1996 (first entry)
 DE Human placental calcium sensor protein.
 KW Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;
 KW membranous glomerulonephritis; therapy; ulcer; wound; placenta.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 1287 /note= "unidentified amino acid"
 FT misc_difference 4094 /note= "unidentified amino acid"
 FT misc_difference 4210 /note= "unidentified amino acid"
 FT WO9615801-A1.
 PD 30-MAY-1996.
 PF 22-NOV-1995; U15203.
 PR 23-NOV-1994; US-344836.
 PR 07-JUN-1995; US-487314.
 PA (RHON) RHONE POULENC RORER PHARM INC.
 PI Amerstrom G, Crumley GR, Hjalim G, Juhlin C, Morse CC;
 PI Murray EM, Rask L;
 DR WPI: 96-268334/27.
 DR N-PSDB: T28774.
 PT Isolated human calcium sensor protein - useful develop prods. for
 PT diagnosis and treatment of conditions, such as membranous
 PT glomerulonephritis
 PS Claim 7; Page 110-128; 236pp; English.
 CC 500 kDa human calcium sensor proteins (CSP) (R97209-11) were
 CC identified as the products of cDNA clones (T28774-76) respectively
 CC obtd. from human placental, kidney and parathyroid cDNA libraries.
 CC The CSPs show homology to LDL receptor proteins. CSP can be isolated
 CC from parathyroid, placental or kidney tubule cells or obtd. by
 CC expression of the encoding sequences in transformed host cells. It is
 CC useful for identifying potential (ant)agonists and in the diagnosis and
 CC treatment of e.g. membranous glomerulonephritis, and also in the
 CC treatment of wounds and ulcers.
 SQ Sequence 4655 AA;

Query Match 50.0%; Score 62; DB 17; Length 4655;
 Best Local Similarity 77.8%; Pred. No. 2.42e+01;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 275 esgcrisiiy 283
 : |||||

Qy 482 DQGRCSIIY 490

RESULT 9
 ID W43314 standard; Protein; 4655 AA.
 AC W43314;

DT 08-JUN-1998 (first entry)
 DE Human parathyroid calcium sensor protein.
 KW Calcium sensor; human; glycoprotein 330; gp330; parathyroid;
 KW low density lipoprotein; LDL; membranous glomerulonephritis;
 KW osteoporosis; hyperthyroidism; therapy.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..25 /label= Sig_peptide
 FT Domain 26..4422 /note= "extracellular domain"
 FT Domain 4423..4445 /note= "transmembrane domain"
 FT Domain 4446..4655 /note= "C-terminal domain"
 FT Misc_difference 675 /note= "encoded by Gty"
 FT Misc_difference 1018 /note= "encoded by ACM"
 FT WO9744050-A1.
 PD 27-NOV-1997.
 PF 20-MAY-1997; U08682.
 PR 23-MAY-1996; US-652877.
 PA (RHON) RHONE-POULENC RORER PHARM INC.
 PI Akerstrom G, Crumley GR, Hjalim G, Juhlin C, Morse CC,
 PI Murray EM, Rask L;
 DR WPI: 98-018219/02.
 DR N-PSDB: V05997.
 PT New human calcium sensor protein and vector encoding inhibitor of gp
 PT 330 mediated transport of amyloid beta protein - useful to treat
 PT membranous glomerulonephritis and central nervous system disorders
 PS Claim 7; Page 177-195; 240pp; English.
 CC This sequence comprises a human parathyroid calcium sensor protein
 CC (hcsp), deduced from the nucleotide sequence (see V05996) of a
 CC clone obtained by PCR amplification. Full-length sequences for
 CC human placenta (see W43312) and kidney (see W43313) are also
 CC claimed. hcsp is considered to be the human homologue of rat
 CC Heymann nephritis antigen gp330 on the basis of sequence
 CC similarity. It is proposed that the calcium sensor is a
 CC universal sensor for recognition of variation in extracellular
 CC calcium, and plays a key role in calcium regulation via different
 CC organ systems. hcsp belongs to the low density lipoprotein
 CC superfamily of glycoproteins. It can be expressed in a claimed
 CC host cell using a vector comprising hcsp cDNA. Also claimed are a
 CC method of using hcsp to treat human membranous glomerulonephritis,
 CC and the use of hcsp to screen for agonists and antagonists of hcsp.
 CC Such (ant)agonists are useful in the treatment of conditions
 CC associated with perturbations in parathyroid hormone levels,
 CC vitamin D3 production, oestrogen levels, osteoclast or osteoblast
 CC activity, calcium secretion and calcium ion homeostasis, e.g.
 CC hyperthyroidism and osteoporosis.
 SQ Sequence 4655 AA;

Query Match 50.0%; Score 62; DB 29; Length 4655;
 Best Local Similarity 77.8%; Pred. No. 2.42e+01;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 275 esgcrisiiy 283
 : |||||

Qy 482 DQGRCSIIY 490

RESULT 10
 ID R97211 standard; Protein; 4655 AA.
 AC R97211;
 DT 31-JUL-1996 (first entry)
 DE Human parathyroid calcium sensor protein.
 KW Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;
 KW membranous glomerulonephritis; therapy; ulcer; wound; parathyroid.
 OS Homo sapiens.
 PN WO9615801-A1.
 PD 30-MAY-1996.
 PF 22-NOV-1995; U15203.

PR 23-NOV-1994; US-344836.
 PR 07-JUN-1995; US-487314.
 PA (RHON) RHONE POULENC RORER PHARM INC.
 PI Amerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC;
 PI Murray EM, Rask L;
 DR WPI: 96-268334/27.
 DR N-PSDB: T28776.
 PT Isolated human calcium sensor protein - useful develop prods. for
 PT diagnosis and treatment of conditions, such as membranous
 PT glomerulonephritis
 PS Claim 7; Page 175-193; 236pp; English.
 CC 500 kDa human calcium sensor proteins (CSP) (R97209-11) were
 CC identified as the products of cDNA clones (T28774-76) respectively
 CC obtd from human placental, kidney and parathyroid cDNA libraries.
 CC The CSPs show homology to LDL receptor proteins. CSP can be isolated
 CC from parathyroid, placental or kidney tubule cells or obtd. by
 CC expression of the encoding sequences in transformed host cells. It is
 CC useful for identifying potential (ant)agonists and in the diagnosis
 CC and treatment of e.g. membranous glomerulonephritis, and also in the
 CC treatment of wounds and ulcers.
 SQ Sequence 4655 AA;
 Query Match 50.0%; Score 62; DB 17; Length 4655;
 Best Local Similarity 77.8%; Pred. No. 2.42e+01;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 275 esgrcisly 283
 : |||||
 QY 482 DQGRCSISY 490
 RESULT 11
 ID R97210 standard; Protein; 4655 AA.
 AC R97210;
 DT 31-JUL-1996 (first entry)
 DE Human kidney calcium sensor protein.
 KW Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;
 KW membranous glomerulonephritis; therapy; ulcer; wound; kidney.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 4210 /note= "unidentified amino acid"
 FT W09615801-A1.
 PN 30-MAY-1996.
 PF 22-NOV-1995; U15203.
 PR 23-NOV-1994; US-344836.
 PR 07-JUN-1995; US-487314.
 PA (RHON) RHONE POULENC RORER PHARM INC.
 PI Amerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC;
 PI Murray EM, Rask L;
 DR WPI: 96-268334/27.
 DR N-PSDB: T28775.
 PT Isolated human calcium sensor protein - useful develop prods. for
 PT diagnosis and treatment of conditions, such as membranous
 PT glomerulonephritis
 PS Claim 7; Page 143-161; 236pp; English.
 CC 500 kDa human calcium sensor proteins (CSP) (R97209-11) were
 CC identified as the products of cDNA clones (T28774-76) respectively
 CC obtd. from human placental, kidney and parathyroid cDNA libraries.
 CC The CSPs show homology to LDL receptor proteins. CSP can be isolated
 CC from parathyroid, placental or kidney tubule cells or obtd. by
 CC expression of the encoding sequences in transformed host cells. It is
 CC useful for identifying potential (ant)agonists and in the diagnosis
 CC and treatment of e.g. membranous glomerulonephritis, and also in the
 CC treatment of wounds and ulcers.
 SQ Sequence 4655 AA;
 Query Match 50.0%; Score 62; DB 17; Length 4655;
 Best Local Similarity 77.8%; Pred. No. 2.42e+01;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 275 esgrcisly 283
 : |||||

QY 482 DQGRCSISY 490
 RESULT 12
 ID P50324 standard; Protein; 339 AA.
 AC P50324;
 DT 20-NOV-1991 (first entry)
 DE Sequence of the last 339 amino acids of
 DE hygromycin B phosphotransferase.
 KW Enzyme; antibiotic resistant marker; cloning vector.
 PN GB2143534-A.
 PD 13-FEB-1985.
 PF 17-JUL-1984; 018171.
 PR 22-JUL-1983; US-516222.
 PR 19-SEP-1983; US-533189.
 PR 26-SEP-1983; US-535508.
 PR 31-MAY-1988; US-205011.
 PA (ELIL) ELI LILLY & CO.
 PI Ingolia TD, Rao RN, Kaster KR;
 DR WPI: 85-040002/07.
 DR N-PSDB: N50387.
 PT Modified hygromycin B resistance conferring gene - useful for
 PT cloning of promoters, construction of gene fusions etc.
 PS Disclosure; Page 28; 39pp; English.
 CC The DNA of the invention (N50387) is useful for cloning, isolation
 CC and characterisation of promoters and for the construction of gene
 CC fusions, acting as dominant markers in host cells. The gene confers
 CC hygromycin B resistance. A DNA segment conferring such resistance is
 CC described in GB2100738. With the vectors and transformants, cloning
 CC of genes directly or indirectly encoding specific functional
 CC polypeptides or fused gene prods. is possible. The DNA is esp. the
 CC 1.3bp Bam HI-Bgl II restriction fragment of plasmid pIT 123. The
 CC vector may be plasmid pIT 123,125,144,208,212,213,215,217 or 219 or
 CC pKC 307 or 308. The transformant is esp. Escherichia coli or
 CC Saccharomyces cerevisiae.
 SQ Sequence 339 AA;
 Query Match 48.4%; Score 60; DB 3; Length 339;
 Best Local Similarity 60.0%; Pred. No. 3.89e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 285 awagrcdai 294
 : | |||| : |
 QY 480 GWDQGRCSISY 489
 RESULT 13
 ID P80056 standard; Protein; 341 AA.
 AC P80056;
 DT 28-JAN-1993 (first entry)
 DE Sequence encoded by the 1.5kb BglII/SacI fragment of plasmid
 DE pKC222 which comprises the hygromycin B resistance gene and
 DE control element.
 DE Aminoglycoside antibiotic; hygromycin B; antibiotic resistance;
 KW marker.
 KW Escherichia coli plasmid pKC203.
 OS US4727028-A.
 PN 23-FEB-1988.
 PD 30-SEP-1983; 538051.
 PR 22-JUN-1981; US-278445.
 PR 26-MAR-1982; US-362215.
 PR 22-APR-1983; US-487787.
 PR 30-SEP-1983; US-538051.
 PA (ELIL) ELI LILLY & CO.
 PI Santerre RF, Rao RN;
 DR WPI: 88-077380/11.
 DR N-PSDB: N80053.
 PT New recombinant DNA cloning vectors - conferring hygromycin B and
 PT G418 resistance to eukaryotic and prokaryotic host cells
 PS Disclosure; Page 5-12; 39pp; English.
 CC Plasmid pKC203 is approximately 15kb in size and can be isolated
 CC readily from E. coli JR225 by conventional procedures. Strain E.
 CC coli JR225 is deposited under ATCC 31912. The 7.5kb BglII

```

CC restriction fragment of pKC203 comprises the structural genes and
CC control elements for expression of resistance to antibiotics
CC hygromycin B and G418. The hygromycin B and G418 resistance genes
CC and control elements in illustrative plasmids pKC214, pKC215, pGD1,
CC pGD2, pGD3 and pGD4, comprise the 7.5 kb BglII fragment of plasmid
CC pKC203. This BglII fragment has been further localized to the
CC 2.5kb Sall/BglII fragment. A restriction map of the 2.5kb
CC Sall/BglII fragment is presented as part of plasmid pKC222.
CC Plasmid pKC222 is useful for isolating the individual genes and
CC control elements that confer resistance to particular antibiotics.
CC For example, the 1.5kb BglII/SacI fragment of plasmid pKC222
CC comprises the hygromycin B resistance gene and control element.
CC Sequence 341 AA;
SQ
Query Match 48.4%; Score 60; DB 5; Length 341;
Best Local Similarity 60.0%; Pred. No. 3.89e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 287 awaaggrcdai 296
QY 480 GWDQGRCSIS 489

RESULT 14
ID R98136 standard; Protein: 524 AA.
AC R98136;
DT 21-OCT-1996 (first entry)
DE DHFR/HPH fusion protein.
KW Expression plasmid; vector: dihydrofolate reductase; DHFR: HPH;
KW hygromycin B phosphotransferase; prothrombin; factor VIII; factor IX;
KW protein C; von Willebrand factor; serum albumin; fusion protein;
KW selection marker; amplification marker; transcription; translation.
OS Synthetic.
PN EP-711835-A1.
PD 15-MAY-1996.
PF 09-NOV-1995; 890202.
PR 14-NOV-1994; AT-002099.
PA (IMMO ) IMMUNO AG.
PI Dörner F, Falkner F, Herlitschka SE, Schlokot U;
DR WPI: 96-232100/24.
PT Expression vector contg. dicistronic unit encoding foreign protein
PT and fusion protein comprising selection and amplification markers -
PT provides high yields of pure foreign protein e.g. blood coagulation
PT factors
PS Claim 32; Page 43-44; 61pp; German.
CC Expression plasmids containing a dicistronic transcription and
CC translation unit including a sequence for a foreign protein and a
CC sequence for a fusion protein comprising a selection marker
CC (hygromycin B phosphotransferase: HPH) and an amplification marker
CC (dihydrofolate reductase; DHFR) can be used for the expression of
CC such foreign proteins. The plasmids provide the foreign protein, for
CC example, prothrombin, factor VIII, factor IX, protein C, von
CC Willebrand factor and/or serum albumin in high yield and purity.
CC They ensure a very high ratio of expressing clones to total clones
CC and a very high initial expression of the foreign protein.
CC Sequence 524 AA;
SQ
Query Match 48.4%; Score 60; DB 18; Length 524;
Best Local Similarity 60.0%; Pred. No. 3.89e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 470 awaaggrcdai 479
QY 480 GWDQGRCSIS 489

RESULT 15
ID R98138 standard; Protein: 539 AA.
AC R98138;
DT 21-OCT-1996 (first entry)
DE DHFR/HPH fusion protein.
KW Expression plasmid; vector: dihydrofolate reductase; DHFR: HPH;
KW hygromycin B phosphotransferase; prothrombin; factor VIII; factor IX;

```

```

KW protein C; von Willebrand factor; serum albumin; fusion protein;
KW selection marker; amplification marker; transcription; translation.
OS Synthetic.
FH Key Location/Qualifiers
FT region 190..194
FT /label= Proline spacer.
PN EP-711835-A1.
PD 15-MAY-1996.
PF 09-NOV-1995; 890202.
PR 14-NOV-1994; AT-002099.
PA (IMMO ) IMMUNO AG.
PI Dörner F, Falkner F, Herlitschka SE, Schlokot U;
DR WPI: 96-232100/24.
DR N-PSDB; T30010.
PT Expression vector contg. di-cistronic unit encoding foreign protein
PT and fusion protein comprising selection and amplification markers -
PT provides high yields of pure foreign protein e.g. blood coagulation
PT factors
PS Claim 32; Page 47-48; 61pp; German.
CC Expression plasmids containing a dicistronic transcription and
CC translation unit including a sequence for a foreign protein and a
CC sequence for a fusion protein comprising a selection marker
CC (hygromycin B phosphotransferase: HPH) and an amplification marker
CC (dihydrofolate reductase; DHFR) can be used for the expression of
CC such foreign proteins. The plasmids provide the foreign protein, for
CC example, prothrombin, factor VIII, factor IX, protein C, von
CC Willebrand factor and/or serum albumin in high yield and purity.
CC They ensure a very high ratio of expressing clones to total clones
CC and a very high initial expression of the foreign protein. This
CC sequence also comprises a poly-Pro spacer region between the DHFR and
CC HPH sequences.
CC Sequence 539 AA;
SQ
Query Match 48.4%; Score 60; DB 18; Length 539;
Best Local Similarity 60.0%; Pred. No. 3.89e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 485 awaaggrcdai 494
QY 480 GWDQGRCSIS 489

```

Search completed: Thu Jul 8 19:01:12 1999
Job time : 24 secs.

ORGANISM #formal_name Emericella nidulans, Aspergillus nidulans
DATE 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
12-Feb-1999
ACCESSIONS S28353
REFERENCE S28353
#authors Mayorga, M.E.; Timberlake, W.E.
#journal Mol. Gen. Genet. (1992) 235:205-212
#title The developmentally regulated Aspergillus nidulans wa gene
encodes a polypeptide homologous to polyketide and fatty
acid synthases.
#cross-references MUID:93101122
#accession S28353
#molecule_type DNA
#residues 1-1986 ##label MAY
##cross-references EMBL:X65866; NID:g5508; PID:g5509

GENETICS
#gene wa
#introns 96/2; 193/3; 1336/3; 1588/3
CLASSIFICATION #superfamily 3-oxoacyl-[acyl-carrier-protein] synthase I
homology; acyl carrier protein homology;
[acyl-carrier-protein] S-malonyltransferase homology
FEATURE
397-805 #domain 3-oxoacyl-[acyl-carrier-protein] synthase I
homology #label OAS\
911-1199 #domain [acyl-carrier-protein] S-malonyltransferase
homology #label AMT\
1648-1718 #domain acyl carrier protein homology #label ACP\
1766-1840 #domain acyl carrier protein homology #label ACP1
SUMMARY #length 1986 #molecular-weight 216633 #checksum 5534
Query Match 50.0%; Score 62; DB 2; Length 1986;
Best Local Similarity 42.9%; Pred. No. 9.53e-01;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 1538 GWDSPCLKFSAD 1551
||| ||| :|||
QY 480 GWDQGRCSIYSSE 493

RESULT 3
ENTRY #type complete
TITLE hygromycin-B kinase (EC 2.7.1.119) - Escherichia coli
plasmids
ORGANISM #formal_name Escherichia coli
DATE 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change
20-Mar-1998
ACCESSIONS A00668; A04603
REFERENCE A00668
#authors Kaster, K.R.; Burgett, S.G.; Rao, R.N.; Ingolia, T.D.
#journal Nucleic Acids Res. (1983) 11:6895-6911
#title Analysis of a bacterial hygromycin B resistance gene by
transcriptional and translational fusions and by DNA
sequencing.
#cross-references MUID:84041492
#accession A00668
#molecule_type DNA
#residues 1-341 ##label KAS
#cross-references GB:V01499; NID:g40922; PID:g40923
##note sequence encoded by Plasmid pKC222, which is derived
from naturally occurring plasmid pKC203

REFERENCE A04603
#authors Gritz, L.; Davies, J.
#journal Gene (1983) 25:179-188
#title Plasmid-encoded hygromycin B resistance: the sequence of
hygromycin B phosphotransferase gene and its expression in
Escherichia coli and Saccharomyces cerevisiae.
#cross-references MUID:84109554
#accession A04603
#molecule_type DNA
#residues 1-341 ##label GRI
##note sequence encoded by Plasmid pJR225

COMMENT This phosphotransferase modifies hygromycin B, an aminoglycoside
antibiotic produced by Streptomyces hygroscopicus.

GENETICS
#gene hph
#genome plasmid
CLASSIFICATION #superfamily hygromycin B phosphotransferase
KEYWORDS antibiotic resistance; phosphotransferase
SUMMARY #length 341 #molecular-weight 38032 #checksum 2442
Query Match 48.4%; Score 60; DB 1; Length 341;
Best Local Similarity 60.0%; Pred. No. 2.24e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 287 AWAQGRDAI 296
: || ||| :|
QY 480 GWDQGRCSI 489

RESULT 4
ENTRY #type complete
TITLE gamma-aminobutyric acid transport protein GAT-1 - Pacific
electric ray
ORGANISM #formal_name Torpedo californica #common_name Pacific
electric ray
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
15-Nov-1996
ACCESSIONS I51368
REFERENCE I51368; S42808
#authors Swanson, G.T.; Umbach, J.A.; Gundersen, C.B.
#journal J. Neurochem. (1994) 63:1-12
#title Glia of the cholinergic electromotor nucleus of Torpedo are
the source of the cDNA encoding a GAT-1-like GABA
transporter.
#cross-references MUID:94267441
#accession I51368
##status preliminary; translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-598 ##label SWA
##cross-references EMBL:X77139; NID:g444009; PID:g444010

GENETICS
#gene TGAT-1
CLASSIFICATION #superfamily gamma-aminobutyric acid transporter
SUMMARY #length 598 #molecular-weight 67219 #checksum 2120
Query Match 48.4%; Score 60; DB 2; Length 598;
Best Local Similarity 50.0%; Pred. No. 2.24e+00;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 166 SWNDRCSNYS 177
: || :|||
QY 480 GWDQGRCSIYS 491

RESULT 5
ENTRY #type complete
TITLE gamma-aminobutyric acid transporter - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
16-Feb-1997
ACCESSIONS F46027
REFERENCE A46027
#authors Liu, Q.R.; Mandiyan, S.; Nelson, H.; Nelson, N.
#journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:6639-6643
#title A family of genes encoding neurotransmitter transporters.
#cross-references MUID:92335351
#accession F46027
##status preliminary
##molecule_type DNA
##residues 1-598 ##label LIU
##experimental_source brain
##note sequence extracted from NCBI backbone (NCBIN:112460,
NCBIP:108818)

CLASSIFICATION #superfamily gamma-aminobutyric acid transporter
KEYWORDS transmembrane protein
SUMMARY #length 598 #molecular-weight 66841 #checksum 7412

Query Match 47.6%; Score 59; DB 2: Length 598;
Best Local Similarity 54.5%; Pred. No. 3.42e+00; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 168 WNTDRCFNSYS 178
l: ||: ||
Qy 481 WDQGRCSIYS 491

RESULT 6
ENTRY S11073 #type complete
TITLE gamma-aminobutyric acid transport protein - human
ALTERNATE_NAMES GABA transporter protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 08-Sep-1997

ACCESSIONS S11073
REFERENCE S11073
#authors Nelson, H.; Mandiyan, S.; Nelson, N.
#journal FEBS Lett. (1990) 269:181-184
#title Cloning of the human brain GABA transporter.
#cross-references MUID:90353567
#accession S11073
#status preliminary
#molecule_type mRNA
#residues 1-599 ##label NEL
#cross-references EMBL:X54673; NID:g31657; PID:g31658
CLASSIFICATION #superfamily gamma-aminobutyric acid transporter
KEYWORDS transmembrane protein
SUMMARY #length 599 #molecular-weight 67014 #checksum 150

Query Match 47.6%; Score 59; DB 2: Length 599;
Best Local Similarity 54.5%; Pred. No. 3.42e+00; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 168 WNTDRCFNSYS 178
l: ||: ||: ||
Qy 481 WDQGRCSIYS 491

RESULT 7
ENTRY ACRTGT #type complete
TITLE gamma-aminobutyric acid transport protein - rat
ALTERNATE_NAMES GABA transporter
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 05-Sep-1997

ACCESSIONS A35918; B35918
REFERENCE A35918
#authors Guastella, J.; Nelson, N.; Nelson, H.; Czyzyk, L.; Keynan, S.; Miedel, M.C.; Davidson, N.; Lester, H.A.; Kanner, B.I.
#journal Science (1990) 249:1303-1306
#title Cloning and expression of a rat brain GABA transporter.
#cross-references MUID:90378307
#accession A35918
#molecule_type mRNA
#residues 1-599 ##label GUA
#cross-references GB:M59742; NID:g204221; PID:g204222
#accession B35918

#molecule_type protein
#residues 546-551; 556-562; 564-566; 572-583 ##label GU2
CLASSIFICATION #superfamily gamma-aminobutyric acid transporter
KEYWORDS brain; glycoprotein; phosphoprotein; transmembrane protein
FEATURE
53-73 #domain transmembrane #status predicted #label TM1\
80-100 #domain transmembrane #status predicted #label TM2\
124-144 #domain transmembrane #status predicted #label TM3\
212-230 #domain transmembrane #status predicted #label TM4\
239-256 #domain transmembrane #status predicted #label TM5\
292-312 #domain transmembrane #status predicted #label TM6\
322-342 #domain transmembrane #status predicted #label TM7\
375-394 #domain transmembrane #status predicted #label TM8\

427-447 #domain transmembrane #status predicted #label TM9\
457-477 #domain transmembrane #status predicted #label TM10\
498-517 #domain transmembrane #status predicted #label TM11\
536-556 #domain transmembrane #status predicted #label TM12\
24,562 #binding_site phosphate (Ser) (covalent) #status predicted\
46 #binding_site phosphate (Thr) (covalent) #status predicted\
176,181,184 #binding_site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 599 #molecular-weight 67001 #checksum 2695

Query Match 47.6%; Score 59; DB 1: Length 599;
Best Local Similarity 54.5%; Pred. No. 3.42e+00; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 168 WNTDRCFNSYS 178
l: ||: ||
Qy 481 WDQGRCSIYS 491

RESULT 8
ENTRY B64514 #type complete
TITLE hypothetical protein MJEC135 - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996

ACCESSIONS B64514
REFERENCE A64300
#authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Claycon, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glöck, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Science (1996) 273:1058-1073
#title Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
#cross-references MUID:96337999
#accession B64514
#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA
#residues 1-630 ##label BUL
#cross-references GB:L77118; NID:gl500644; TIGR:MJEC135; PID:gl522669

GENETICS
#map_position ECLREV45004-43112
#genome plasmid
#note this stable 58-kilobase pair plasmid is also designated ECL (large extrachromosomal element) and contains 44 predicted coding regions

SUMMARY #length 630 #molecular-weight 72440 #checksum 9196

Query Match 47.6%; Score 59; DB 2: Length 630;
Best Local Similarity 63.6%; Pred. No. 3.42e+00; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 588 GWDSGRSERY 598
l: ||: ||: ||
Qy 480 GWDQGRCSIYS 490

RESULT 9
ENTRY B70324 #type complete
TITLE hypothetical protein ag_264 - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998

```

ACCESSIONS B70324
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
          Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;
          Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
          J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
          Aquifex aeolicus.
#cross-references MUID:98196666
#accession B70324
#status preliminary; nucleic acid sequence not shown;
          translation not shown
#molecule_type DNA
#residues 1-217 #label AQF
#cross-references GB:AE000681; NID:g2982963; PID:g2982970; GB:AE000657
#experimental_source strain VF5

GENETICS
#gene aq_264
#summary #length 217 #molecular-weight 25616 #checksum 498

Query Match 46.8%; Score 58; DB 2; Length 217;
Best Local Similarity 43.8%; Pred.No. 5.18e+00;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Db 72 GWDLSFCIVDYKNGRN 87
   ||| : || : || :
Qy 480 GWDGRCISYSSERS 495

RESULT 10
ENTRY B40016 #type fragment
TITLE matrin 3 - human (fragment)
ORGANISM #formal_name Homo Sapiens #common_name man
DATE 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change
06-Dec-1996
ACCESSIONS B40016
REFERENCE A40016
#authors Belgrader, P.; Dey, R.; Berezney, R.
#journal J. Biol. Chem. (1991) 266:9893-9899
#title Molecular cloning of matrin 3. A 125-kilodalton protein of
          the nuclear matrix contains an extensive acidic domain.
#cross-references MUID:91236771
#accession B40016
#status preliminary
#molecule_type mRNA
#residues 1-417 #label BEL
#cross-references GB:M63483
#keywords nuclear matrix
#summary #length 417 #checksum 6905

Query Match 46.8%; Score 58; DB 2; Length 417;
Best Local Similarity 50.0%; Pred.No. 5.18e+00;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 128 WFGQRCVKVDLSEK 141
   ||| : || : || :
Qy 481 WDGRCISYSSER 494

RESULT 11
ENTRY A40016 #type complete
TITLE matrin 3 - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change
06-Dec-1996
ACCESSIONS A40016
REFERENCE A40016
#authors Belgrader, P.; Dey, R.; Berezney, R.
#journal J. Biol. Chem. (1991) 266:9893-9899
#title Molecular cloning of matrin 3. A 125-kilodalton protein of
          the nuclear matrix contains an extensive acidic domain.
#cross-references MUID:91236771

```

```

#accession A40016
#status preliminary
#molecule_type mRNA
#residues 1-845 #label BEL
#cross-references GB:M63485
#keywords nuclear matrix
#summary #length 845 #molecular-weight 95524 #checksum 8549

Query Match 46.8%; Score 58; DB 2; Length 845;
Best Local Similarity 50.0%; Pred.No. 5.18e+00;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 558 WFGQRCVKVDLSEK 571
   ||| : || : || :
Qy 481 WDGRCISYSSER 494

RESULT 12
ENTRY JN0804 #type complete
TITLE nitrate reductase (NADPH) (EC 1.6.6.3) - smut fungus
          (Ustilago maydis)
ORGANISM #formal_name Ustilago maydis #common_name corn smut
DATE 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change
17-Mar-1999
ACCESSIONS JN0804; S25516
REFERENCE JN0804
#authors Banks, G.R.; Shelton, P.A.; Kanuga, N.; Holden, D.W.; Spanos,
          A.
#journal Gene (1993) 131:69-78
#title The Ustilago maydis nar1 gene encoding nitrate reductase
          activity: Sequence and transcriptional regulation.
#cross-references MUID:93380675
#accession JN0804
#molecule_type DNA
#residues 1-908 #label BAN
#cross-references EMBL:X67687; NID:g5226; PID:g5227

GENETICS
#gene nar1
#classification #superfamily nitrate reductase (NADH); cytochrome b5 core
          homology; cytochrome-b5 reductase homology;
          molybdopterin-binding domain homology
#keywords electron transfer; FAD; flavoprotein; heme; molybdenum; NADP;
          nitrate assimilation; oxidoreductase
#feature 8-435
          #domain molybdopterin-binding domain homology #label
          PCO\
          #domain cytochrome b5 core homology #label CB5\
          #domain cytochrome-b5 reductase homology #label CBR
          #length 908 #molecular-weight 101524 #checksum 9196

Query Match 46.8%; Score 58; DB 2; Length 908;
Best Local Similarity 45.5%; Pred.No. 5.18e+00;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 898 GWDLQRCVVF 908
   ||| : || : || :
Qy 480 GWDGRCISY 490

RESULT 13
ENTRY T00021 #type complete
TITLE DN-cadherin - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
22-Jan-1999
ACCESSIONS T00021
REFERENCE Z14058
#authors Iwai, Y.; Usui, T.; Hirano, S.; Steward, R.; Takeichi, M.;
          Uemura, T.
#journal Neuron (1997) 19:77-89
#title Axon patterning requires DN-cadherin, a novel neuronal
          adhesion receptor, in the Drosophila embryonic CNS.
#accession T00021

```

##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-3097 ##label IWA
##cross-references EMBL:AB002397; NID:d1134177; PID:d1023011
##experimental_source whole embryo and adult head

GENETICS

##map_position 36D
##note 1(2)36Da
SUMMARY #length 3097 #molecular-weight 347189 #checksum 6232

Query Match 46.8%; Score 58; DB 2; Length 3097;
Best Local Similarity 35.7%; Pred. No. 5.18e+00;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 2602 WEHGNCGSLSEAR 2615

OY 481 WDGRCSISYSSER 494
|::|::|
|::|::|

RESULT 14
ENTRY C54252 #type complete
TITLE omega-agatoxin III, 8.607K - funnel-weaving spider
(Agelenopsis aperta)
ORGANISM #formal_name Agelenopsis aperta
DATE 06-Oct-1994 #sequence_revision 18-Nov-1994 #text_change 19-Jan-1996
ACCESSIONS C54252
REFERENCE A54252
#authors Ertel, E.A.; Warren, V.A.; Adams, M.E.; Griffin, P.R.; Cohen, C.J.; Smith, M.M.
#journal Biochemistry (1994) 33:5098-5108
#title Type III omega-agatoxins: a family of probes for similar binding sites on L- and N-type calcium channels.
#cross-references MUID:94227039
#accession C54252
##status preliminary
##molecule_type protein
##residues 1-76 ##label ERT
##experimental_source venom
##note sequence extracted from NCBI backbone (NCBIP:148241)
SUMMARY #length 76 #molecular-weight 8620 #checksum 8933

Query Match 46.0%; Score 57; DB 2; Length 76;
Best Local Similarity 38.5%; Pred. No. 7.81e+00;
Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 63 DPDKCVSVYKPKR 75

OY 482 DOGRCSISYSSER 494
|::|::|
|::|::|

RESULT 15
ENTRY B54252 #type complete
TITLE omega-agatoxin III, 8.637K - funnel-weaving spider
(Agelenopsis aperta)
ORGANISM #formal_name Agelenopsis aperta
DATE 06-Oct-1994 #sequence_revision 18-Nov-1994 #text_change 19-Jan-1996
ACCESSIONS B54252
REFERENCE A54252
#authors Ertel, E.A.; Warren, V.A.; Adams, M.E.; Griffin, P.R.; Cohen, C.J.; Smith, M.M.
#journal Biochemistry (1994) 33:5098-5108
#title Type III omega-agatoxins: a family of probes for similar binding sites on L- and N-type calcium channels.
#cross-references MUID:94227039
#accession B54252
##status preliminary
##molecule_type protein
##residues 1-76 ##label ERT
##experimental_source venom
##note sequence extracted from NCBI backbone (NCBIP:148240)
SUMMARY #length 76 #molecular-weight 8648 #checksum 9178

Query Match 46.0%; Score 57; DB 2; Length 76;
Best Local Similarity 38.5%; Pred. No. 7.81e+00;
Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 63 DPDKCVSVYKPKR 75

OY 482 DOGRCSISYSSER 494
|::|::|
|::|::|

Search completed: Thu Jul 8 19:00:28 1999
Job time : 14 secs.

THIS PAGE BLANK (USPTO)

WQSERELH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:59:16 1999; MasPar time 3.67 Seconds

Tabular output not generated. 123.145 Million cell updates/sec

Title: >US-09-041-236-2
Description: (480-495) from US09041236.pep (26 of 45)
Perfect Score: 124
Sequence: 1 GWDGRCISYSSRS 16

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 27.660; Variance 34.099; scale 0.811

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	66	53.2	479	1	OPRJ_PSEAE	OUTER MEMBRANE PROTEIN	4.17e+02
2	65	52.4	678	1	SCOB_EMENI	SULFUR METABOLITE REPR	6.82e-02
3	64	51.6	831	1	SYFB_MYCTU	PHENYLALANYL-TRNA SYNT	1.11e-01
4	62	50.0	473	1	YS81_CAEEL	HYPOTHETICAL 52.5 KD P	2.91e-01
5	62	50.0	1986	1	WA_EMENI	CONIDIAL GREEN PIGMENT	2.91e-01
6	62	50.0	4655	1	LRP2_HUMAN	LOW-DENSITY LIPOPROTEIN	2.91e-01
7	60	48.4	341	1	KHYB_ECOLI	HYGROMYCIN-B KINASE (E	7.47e-01
8	59	47.6	598	1	NTG1_MUSCO	SODIUM- AND CHLORIDE-D	1.19e+00
9	59	47.6	599	1	NTG1_MOUSE	SODIUM- AND CHLORIDE-D	1.19e+00
10	59	47.6	599	1	NTG1_RAT	SODIUM- AND CHLORIDE-D	1.19e+00
11	59	47.6	599	1	NTG1_HUMAN	SODIUM- AND CHLORIDE-D	1.19e+00
12	59	47.6	630	1	YZ35_METJA	HYPOTHETICAL PROTEIN M	1.19e+00
13	58	46.8	419	1	MAT3_HUMAN	MATRIN 3 (FRAGMENT)	1.87e+00
14	58	46.8	845	1	MAT3_RAT	MATRIN 3	1.87e+00
15	58	46.8	908	1	NIA_USTWA	NITRATE REDUCTASE (NAD	1.87e+00
16	57	46.0	183	1	VCO7_ADEB2	MAJOR CORE PROTEIN PRE	2.94e+00
17	57	46.0	329	1	E2B1_METJA	PUTATIVE TRANSLATION I	2.94e+00
18	57	46.0	524	1	G6PD_SYNP7	GLUCOSE-6-PHOSPHATE 1-	2.94e+00
19	57	46.0	684	1	CC4_CANAL	CELL DIVISION CONTROL	2.94e+00
20	57	46.0	1403	1	NKCR_HUMAN	NK-TUMOR RECOGNITION P	2.94e+00
21	57	46.0	1560	1	X169_HUMAN	XE169 PROTEIN (SMCX PR	2.94e+00
22	56	45.2	141	1	FKB2_HUMAN	FK506-BINDING PROTEIN	4.59e+00
23	56	45.2	238	1	RNRH_RHINI	RIBONUCLEASE RH PRECUR	4.59e+00

24	56	45.2	240	1	LECS_VATMA	SEED LECTIN (VML).	4.59e+00
25	56	45.2	430	1	XQTV_BACSU	HYPOTHETICAL 44.7 KD P	4.59e+00
26	56	45.2	1640	1	CO3_ONCMY	COMPLEMENT C3-1 (CONTA	4.59e+00
27	56	45.2	4753	1	LRP_CAEEL	LOW-DENSITY LIPOPROTEI	4.59e+00
28	55	44.4	152	1	S-ADENOSYL-L-METHIONIN	7.13e+00	
29	55	44.4	180	1	X169_CRIGR	XE169 PROTEIN (FRAGMEN	7.13e+00
30	55	44.4	322	1	ACSD_MOOTH	CORRINOID/IRON-SULFUR	7.13e+00
31	55	44.4	341	1	AAPJ_RHLV	GENERAL L-AMINO ACID-B	7.13e+00
32	55	44.4	426	1	TWIN_DROME	CBC25-LIKE PROTEIN PHO	7.13e+00
33	55	44.4	585	1	CO8A_RABIT	COMPLEMENT COMPONENT C	7.13e+00
34	55	44.4	720	1	YD71_SCHPO	HYPOTHETICAL 82.5 KD P	7.13e+00
35	55	44.4	1213	1	X169_MOUSE	XE169 PROTEIN (SMCX PR	7.13e+00
36	55	44.4	1274	1	SRF3_BACSU	SURFACTIN SYNTHETASE S	7.13e+00
37	54	43.5	80	1	YTM8_BACSU	HYPOTHETICAL 9.3 KD PR	1.10e+01
38	54	43.5	318	1	RECA_BACFR	RECA PROTEIN.	1.10e+01
39	54	43.5	340	1	RECA_PORGI	SERINE/THREONINE PROTE	1.10e+01
40	54	43.5	499	1	PPP5_HUMAN	SERINE/THREONINE PROTE	1.10e+01
41	54	43.5	459	1	YHV9_YEAST	HYPOTHETICAL 56.3 KD P	1.10e+01
42	54	43.5	504	1	EPAY_RAT	EPHRIN TYPE-A RECEPTOR	1.10e+01
43	54	43.5	998	1	EPAY_HUMAN	EPHRIN TYPE-A RECEPTOR	1.10e+01
44	54	43.5	998	1	EPAY_MOUSE	EPHRIN TYPE-A RECEPTOR	1.10e+01
45	54	43.5	998	1	EPAY_MOUSE	EPHRIN TYPE-A RECEPTOR	1.10e+01

ALIGNMENTS

RESULT 1	OPRJ_PSEAE	STANDARD;	PRT;	479 AA.
ID	Q51397;			
AC	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	OUTER MEMBRANE PROTEIN OPRJ PRECURSOR.			
GN	OPRJ.			
OS	PSEUDOMONAS AERUGINOSA.			
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;			
OC	PSEUDOMONAS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PAO;			
RA	POOLE K., GOTOH N., TSUJIMOTO H., ZHAO Q., WADA A., YAMASAKI T.,			
RA	NESHAT S., YAMAGISHI J., NISHINO T.;			
RL	SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.			
CC	-!- FUNCTION: CHANNEL-FORMING COMPONENT OF A MULTIDRUG RESISTANCE			
CC	EFFLUX PUMP.			
CC	-!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID			
CC	ANCHOR (PROBABLE).			
CC	-!- SIMILARITY: BELONGS TO THE FUSA/NOTD FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; U57969; G1399759; -			
DR	PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.			
KW	OUTER MEMBRANE; ANTIOTIC RESISTANCE; LIPOPROTEIN; SIGNAL.			
FT	SIGNAL 1 19 BY SIMILARITY.			
FT	CHAIN 20 479 OUTER MEMBRANE PROTEIN OPRJ.			
FT	LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).			
SQ	SEQUENCE 479 AA; 51873 MW; AB7ECOAC CRC32;			
Query Match	53.2%;	Score 66;	DB 1;	Length 479;
Best Local Similarity	43.8%;	Pred. No. 4.17e-02;		
Matches	7;	Conservative	4;	Mismatches 5; Indels 0; Gaps 0;
Db	464 GWDGRCISYSSRS 479			
QY	480 GWDGRCISYSSRS 495			

```

RA DAVIES R., DEVLIN K., FELTWELL T., GENTILES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL NATURE 393:537-544(1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) --> AMP +
CC PYROPHOSPHATE + L-PHENYLALANYL-TRNA(PHE).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z85982; E300633; -.
DR HSSP; P27002; 1PYS.
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
SQ SEQUENCE 831 AA; 88373 MW; BFB474AA CRC32;

Query Match 51.6%; Score 64; DB 1; Length 831;
Best Local Similarity 40.0%; Pred. No. 1.11e-01;
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Db 680 WHPGRCAQVFGVSS 694
| | | | | : | |
Qy 481 WDQGRCSISYSSRS 495

RESULT 4
ID YS81_CAEEL STANDARD; PRT; 473 AA.
AC Q09621;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 52.5 KD PROTEIN ZK945.1 IN CHROMOSOME II.
GN ZK945.1.
OS CAENORHABDITIS ELEGANS
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WILKINSON-SPROAT J.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48544; E1351785; -.
DR WORMPEP; ZK945.1; CE01731.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 26 48 POTENTIAL.
FT TRANSMEM 125 143 POTENTIAL.
SQ SEQUENCE 473 AA; 52546 MW; 3D5A38DE CRC32;

Query Match 50.0%; Score 62; DB 1; Length 473;
Best Local Similarity 37.5%; Pred. No. 2.91e-01;
Matches 6; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 86 GWREGASITVYHKDR 101

```

```

RA DAVIES R., DEVLIN K., FELTWELL T., GENTILES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL NATURE 393:537-544(1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) --> AMP +
CC PYROPHOSPHATE + L-PHENYLALANYL-TRNA(PHE).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z85982; E300633; -.
DR HSSP; P27002; 1PYS.
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
SQ SEQUENCE 831 AA; 88373 MW; BFB474AA CRC32;

Query Match 51.6%; Score 64; DB 1; Length 831;
Best Local Similarity 40.0%; Pred. No. 1.11e-01;
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Db 680 WHPGRCAQVFGVSS 694
| | | | | : | |
Qy 481 WDQGRCSISYSSRS 495

RESULT 4
ID YS81_CAEEL STANDARD; PRT; 473 AA.
AC Q09621;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 52.5 KD PROTEIN ZK945.1 IN CHROMOSOME II.
GN ZK945.1.
OS CAENORHABDITIS ELEGANS
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WILKINSON-SPROAT J.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48544; E1351785; -.
DR WORMPEP; ZK945.1; CE01731.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 26 48 POTENTIAL.
FT TRANSMEM 125 143 POTENTIAL.
SQ SEQUENCE 473 AA; 52546 MW; 3D5A38DE CRC32;

Query Match 50.0%; Score 62; DB 1; Length 473;
Best Local Similarity 37.5%; Pred. No. 2.91e-01;
Matches 6; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 86 GWREGASITVYHKDR 101

```

```
QY 480 GWD-QGRCSISYSSR 494
||: || ||: ||: ||
RESULT 5
ID WA_EMENI STANDARD: PRT: 1986 AA.
AC Q03149;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CONIDIAL GREEN PIGMENT SYNTHASE (EC 2.3.1.-).
GN WA.
OS EMERICELLA NIDULANS (ASPERGILLUS NIDULANS).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
OC EUROTIALES; TRICHOCOMACEAE; EMERICELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93101122.
RA MAYORGA M.E., TIMBERLAKE W.E.;
RT "The developmentally regulated Aspergillus nidulans wa gene encodes a
RT polypeptide homologous to polyketide and fatty acid synthases.";
RL MOL. GEN. GENET. 235:205-212(1992).
CC -!- FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN
CC INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED
CC BY CONIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE
CC ASEXUAL SPORES (CONIDIA).
CC -!- COFACTOR: CONTAINS TWO COVALENTLY BOUND PHOSPHOPANTHETINES
CC (POTENTIAL).
CC -!- PATHWAY: BIOSYNTHESIS OF CONIDIAL GREEN PIGMENT.
CC -!- SIMILARITY: WITH BOTH EUKARYOTIC AND PROKARYOTIC POLYKETIDE
CC SYNTHASES AND VERTEBRATE FATTY ACID SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X65866; G5509; -
DR PIR: S28353; S28353.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS50075; ACP_DOMAIN; 2.
DR PFAM: PF00109; ketoacyl-synt; 1.
DR PFAM: PF00550; pp-binding; 2.
DR PFAM: PF00698; Acyl transf; 1.
KW TRANSFERASE; PHOSPHOPANTHETHEINE; MULTIFUNCTIONAL ENZYME.
FT DOMAIN 529 582 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT DOMAIN 991 1024 ACYL/MALONYL TRANSFERASES
FT DOMAIN 1650 1719 (BY SIMILARITY).
FT DOMAIN 1772 1841 ACYL CARRIER (ACP).
FT ACT_SITE 548 548 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 1001 1001 ACYL/MALONYL TRANSFERASES (BY
FT SIMILARITY).
FT BINDING 1682 1682 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 1804 1804 PHOSPHOPANTHETHEINE (BY SIMILARITY).
SQ SEQUENCE 1986 AA; 216634 MW; 2EAD2D73 CRC32;

Query Match 50.0%; Score 62; DB 1; Length 1986;
Best Local Similarity 42.9%; Pred. No. 2,91e-01;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 1538 GWDMSRCLKKFSAD 1551
||| ||: ||: ||
QY 480 GWDQGRCSISYSSE 493
RESULT 6
ID LRP2_HUMAN STANDARD: PRT: 4655 AA.
AC P98164; O00711; Q16215;
```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
-----
DR EMBL; U33837; G1809240; -
DR EMBL; U04441; G1389559; -
DR EMBL; S73145; G685061; -
DR MTM; 600073; -
DR PROSITE; PS00010; ASX-HYDROXYL; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 9.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS01209; LDLRA_1; 31.
DR PROSITE; PS50068; LDLRA_2; 36.
DR PFAM; PF00037; ldl_recept_a; 36.
DR PFAM; PF00058; ldl_recept_b; 37.
DR HSP; P01130; IAJJ.
KW GLYCOPROTEIN; REPEAT; ENDOCYTOSIS; COATED PITS; TRANSMEMBRANE;
KW RECEPTOR; EGF-LIKE DOMAIN; SIGNAL; POLYMORPHISM.
FT SIGNAL 1 25
FT CHAIN 26 4655
FT DOMAIN 26 4423
FT TRANSMEM 4424 4446
FT DOMAIN 4447 4655
FT DOMAIN 26 64
FT DOMAIN 65 105
FT DOMAIN 106 144
FT DOMAIN 145 181
FT DOMAIN 182 219
FT DOMAIN 220 258
FT DOMAIN 264 308
FT DOMAIN 309 347
FT DOMAIN 348 386
FT DOMAIN 436 477
FT DOMAIN 479 520
FT DOMAIN 522 567
FT DOMAIN 569 612
FT DOMAIN 613 653
FT DOMAIN 659 705
FT DOMAIN 753 794
FT DOMAIN 796 836
FT DOMAIN 838 880
FT DOMAIN 882 924
FT DOMAIN 970 1014
FT DOMAIN 1024 1062
FT DOMAIN 1065 1103
FT DOMAIN 1107 1145
FT DOMAIN 1147 1185
FT DOMAIN 1186 1224
FT DOMAIN 1228 1268
FT DOMAIN 1269 1307
FT DOMAIN 1310 1350
FT DOMAIN 1349 1389
FT DOMAIN 1390 1429
FT DOMAIN 1429 1519
FT DOMAIN 1478 1519
FT DOMAIN 1521 1562
FT DOMAIN 1566 1608
FT DOMAIN 1610 1653
FT DOMAIN 1655 1695
FT DOMAIN 1695 1741
FT DOMAIN 1700 1741
FT DOMAIN 1741 1831
FT DOMAIN 1831 1881
FT DOMAIN 1883 1921
FT DOMAIN 1931 1971
FT DOMAIN 1972 2012
FT DOMAIN 2012 2059
FT DOMAIN 2059 2155
FT DOMAIN 2107 2155
FT DOMAIN 2157 2200
FT DOMAIN 2202 2244
FT DOMAIN 2244 2288
FT DOMAIN 2288 2331
FT DOMAIN 2331 2383
FT DOMAIN 2383 2476
FT DOMAIN 2476 2517
FT DOMAIN 2517 2561
FT DOMAIN 2561 2603
FT DOMAIN 2603 2646
FT DOMAIN 2646 2693
FT DOMAIN 2693 2738
FT DOMAIN 2738 2777
FT DOMAIN 2777 2819
FT DOMAIN 2819 2861
FT DOMAIN 2861 2901
FT DOMAIN 2901 2945
FT DOMAIN 2945 2990
FT DOMAIN 2990 3029
FT DOMAIN 3029 3070
FT DOMAIN 3070 3110
FT DOMAIN 3110 3151
FT DOMAIN 3151 3192
FT DOMAIN 3192 3239
FT DOMAIN 3239 3282
FT DOMAIN 3282 3333
FT DOMAIN 3333 3375
FT DOMAIN 3375 3418
FT DOMAIN 3418 3459
FT DOMAIN 3459 3509
FT DOMAIN 3509 3550
FT DOMAIN 3550 3591
FT DOMAIN 3591 3632
FT DOMAIN 3632 3673
FT DOMAIN 3673 3716
FT DOMAIN 3716 3756
FT DOMAIN 3756 3795
FT DOMAIN 3795 3834
FT DOMAIN 3834 3880
FT DOMAIN 3880 3922
FT DOMAIN 3922 3964
FT DOMAIN 3964 4006
FT DOMAIN 4006 4048
FT DOMAIN 4048 4154
FT DOMAIN 4154 4195
FT DOMAIN 4195 4239
FT DOMAIN 4239 4282
FT DOMAIN 4282 4330
FT DOMAIN 4330 4368
FT DOMAIN 4368 4411
FT SITE 1742 1744
FT SITE 1744 1744
FT SITE 4526 4526
FT SITE 4526 4598
FT DISULFID 28 40
FT DISULFID 35 53
FT DISULFID 47 62
FT DISULFID 67 80
FT DISULFID 74 93
...
Note: remainder of annotations omitted.

Query Match 50.0%; Score 62; DB 1; Length 4655;
Best Local Similarity 77.8%; Pred. No. 2.91e-01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 275 ESGRCISY 283
: |||||
QY 482 DQGRCSY 490

RESULT 7
ID KHYB_ECOLI STANDARD; PRT; 341 AA.
AC P00557;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE HYGROMYCIN-B KINASE (EC 2.7.1.119) (HYGROMYCIN B PHOSPHOTRANSFERASE)
DE (APH(7'')).
GN HPH.
```

OS ESCHERICHIA COLI.
 OG PLASMID PKC222, AND PLASMID PJR225.
 CC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 CC ESCHERICHIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12; PLASMID-PKC222;
 RX MEDLINE: 84041492.
 RA KASTER K.R., BURGETT S.G., RAO R.N., INGOLIA T.D.;
 RT "Analysis of a bacterial hygromycin B resistance gene by
 RT transcriptional and translational fusions and by DNA sequencing.";
 RL NUCLEIC ACIDS RES. 11:6895-6911(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC PLASMID-PJR225.
 RX MEDLINE: 84109554.
 RA GRITZ L., DAVIES J.;
 RT "Plasmid-encoded hygrocin B resistance: the sequence of hygromycin
 RT B phosphotransferase gene and its expression in Escherichia coli and
 RT Saccharomyces cerevisiae";
 RL GENE 25:179-188(1983).
 RN [3]
 RP SEQUENCE OF 1-39 FROM N.A.
 RC MEDLINE: 84092826.
 RA BRAEU B., PILZ U., PIEPERSBERG W.;
 RT "Genes for gentamicin-(3)-N-acetyltransferases III and IV: I.
 RT Nucleotide sequence of the AAC(3)-IV gene and possible involvement of
 RT an IS140 element in its expression.";
 RL MOL. GEN. GENET. 193:179-187(1984).
 CC -!- FUNCTION: THE AMINOGLYCOSIDE PHOSPHOTRANSFERASES ACHIEVE
 CC INACTIVATION OF THEIR ANTIBIOTIC SUBSTRATES BY PHOSPHORYLATION.
 CC -!- CATALYTIC ACTIVITY: ATP + HYGROMYCIN B = ADP +
 CC 7'-O-PHOSPHOHYGROMYCIN B.
 CC -!- HYGROMYCIN B RESISTANT E. COLI WERE FOUND TO CONTAIN A KLEBSIELLA
 CC DERIVED PLASMID, PJR225. THIS PLASMID CONTAINS THE HPH GENE.
 CC -!- SIMILARITY: TO OTHER AMINOGLYCOSIDE PHOSPHOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X01193; G150776; -.
 DR EMBL; V01499; G40923; -.
 DR EMBL; X01385; G40877; -.
 DR EMBL; X89856; G953210; -.
 DR EMBL; X89857; G953211; -.
 DR PIR; A00668; WGECH.
 DR PIR; A04603; WGECHJ.
 KW ANTIBIOTIC RESISTANCE; TRANSFERASE; KINASE; ATP-BINDING; PLASMID.
 SQ SEQUENCE 341 AA; 38032 MW; 39115E06 CRC32;
 Query Match 48.4%; Score 60; DB 1; Length 341;
 Best Local Similarity 60.0%; Pred. No. 7.47e-01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 287 AWAQRCDAI 296
 QY 480 GWDQGRCSI 489
 RESULT 8
 ID NTGL MUSCO STANDARD; PRT; 598 AA.
 AC P40057;
 DT 01-FEB-1996 (REL. 33, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE SODIUM- AND CHLORIDE-DEPENDENT GABA TRANSPORTER 1.
 GN SLC6A1 OR GAT1 OR GAT-1.
 OS MUS COOKII.

CC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA; EUTHERIA;
 CC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-RETINA;
 RA RUIZ M., EGAL H., QIAN X.J., SARTHY V.P., SARKAR H.K.;
 RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
 CC -!- FUNCTION: TERMINATES THE ACTION OF GABA BY ITS HIGH AFFINITY
 CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: BRAIN.
 CC -!- THIS PROTEIN IS THE TARGET OF PSYCHOMOTOR STIMULANTS SUCH AS
 CC AMPHETAMINES OR COCAINE.
 CC -!- SIMILARITY: BELONGS TO THE SODIUM-NEUROTRANSMITTER SYMPORTER
 CC FAMILY (SNF).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L32178; G533226; -.
 DR MGD; MGI:95627; GABT1.
 DR PROSITE; PS00610; NA-NEUROTRAN_SYMP_1; 1.
 DR PROSITE; PS00754; NA-NEUROTRAN_SYMP_2; 1.
 DR PFAM; PF00209; SNF; 1.
 KW NEUROTRANSMITTER TRANSPORT; TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN;
 KW SYMPORT; MULTIENE FAMILY.
 FT DOMAIN 1 52 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 53 73 1 (POTENTIAL).
 FT TRANSMEM 81 100 2 (POTENTIAL).
 FT TRANSMEM 124 144 3 (POTENTIAL).
 FT DOMAIN 145 211 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 212 229 4 (POTENTIAL).
 FT TRANSMEM 238 255 5 (POTENTIAL).
 FT TRANSMEM 291 308 6 (POTENTIAL).
 FT TRANSMEM 320 341 7 (POTENTIAL).
 FT TRANSMEM 374 393 8 (POTENTIAL).
 FT TRANSMEM 421 439 9 (POTENTIAL).
 FT TRANSMEM 456 476 10 (POTENTIAL).
 FT TRANSMEM 497 516 11 (POTENTIAL).
 FT TRANSMEM 535 553 12 (POTENTIAL).
 FT DOMAIN 554 598 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 176 176 POTENTIAL.
 FT CARBOHYD 181 181 POTENTIAL.
 FT CARBOHYD 184 184 POTENTIAL.
 SQ SEQUENCE 598 AA; 66783 MW; 5BCE8973 CRC32;
 Query Match 47.6%; Score 59; DB 1; Length 598;
 Best Local Similarity 54.5%; Pred. No. 1.19e+00;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Db 168 WNTDRCSNYS 178
 QY 481 WDGRCISYIS 491
 RESULT 9
 ID NTGL MOUSE STANDARD; PRT; 598 AA.
 AC P31648;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE SODIUM- AND CHLORIDE-DEPENDENT GABA TRANSPORTER 1.
 GN SLC6A1 OR GAT1 OR GAT-1.
 OS MUS MUSCULUS (MOUSE).
 CC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA; EUTHERIA;
 CC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.

```

RX MEDLINE; 92335351.
RA LIU Q.-R., MANDIYAN S., NELSON H., NELSON N.;
RT "A family of genes encoding neurotransmitter transporters.";
RL PROC. NATL. ACAD. SCI. U.S.A. 89:6639-6643(1992).
CC -1- FUNCTION: TERMINATES THE ACTION OF GABA BY ITS HIGH AFFINITY
CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- THIS PROTEIN IS THE TARGET OF PSYCHOMOTOR STIMULANTS SUCH AS
CC AMPHETAMINES OR COCAINE.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M92378; NOT_ANNOTATED_CDS.
DR PIR; F46027; F46027.
DR MGI; MGI:95627; GABT1.
DR PROSITE; PS00610; NA_NEUROTRAN_SYMP_1; 1.
DR PFAM; PF00754; NA_NEUROTRAN_SYMP_2; 1.
DR PFAM; PF00209; SNF; 1.
DR KW NEUROTRANSMITTER TRANSPORT; TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN;
DR SYMPORT; MULTIGENE FAMILY.
FT DOMAIN 1 52 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 53 73
FT TRANSMEM 81 100
FT TRANSMEM 124 144
FT DOMAIN 145 211 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 212 230
FT TRANSMEM 239 256
FT TRANSMEM 292 309
FT TRANSMEM 321 342
FT TRANSMEM 375 394
FT TRANSMEM 422 440
FT TRANSMEM 457 477
FT TRANSMEM 498 517
FT TRANSMEM 536 554
FT TRANSMEM 555 599
FT DOMAIN 176 176 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 181 181 POTENTIAL.
FT CARBOHYD 184 184 POTENTIAL.
FT CARBOHYD 184 184 POTENTIAL.
SQ SEQUENCE 598 AA; 66841 MW; 85D29291 CRC32;

Query Match 47.6%; Score 59; DB 1; Length 598;
Best Local Similarity 54.5%; Pred. No. 1.19e+00;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 168 WNTDRCSFNSYS 178
QY 481 WDOGRCSISYS 491
I: | | | | |

RESULT 10
ID NTGL_RAT STANDARD; PRT; 599 AA.
AC P23978;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SODIUM- AND CHLORIDE-DEPENDENT GABA TRANSPORTER 1.
GN SLC6A1 OR GABT1 OR GAT-1.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE; 90378307.
RC TISSUE-BRAIN.
RX MEDLINE; 90378307.

RA GUASTELLA J., NELSON N., NELSON H., CZYZYK L., KEYNAN S.,
RA MIEDEL M.C., DAVIDSON N., LESTER H.A., KANNER B.I.;
RT "Cloning and expression of a rat brain GABA transporter.";
RL SCIENCE 249:1303-1306(1990).
CC -1- FUNCTION: TERMINATES THE ACTION OF GABA BY ITS HIGH AFFINITY
CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- THIS PROTEIN IS THE TARGET OF PSYCHOMOTOR STIMULANTS SUCH AS
CC AMPHETAMINES OR COCAINE.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59742; G204222; -.
DR PIR; A35918; ACRTGT.
DR PROSITE; PS00610; NA_NEUROTRAN_SYMP_1; 1.
DR PROSITE; PS00754; NA_NEUROTRAN_SYMP_2; 1.
DR PFAM; PF00209; SNF; 1.
DR KW NEUROTRANSMITTER TRANSPORT; TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN;
DR SYMPORT; MULTIGENE FAMILY.
FT DOMAIN 1 52 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 53 73
FT TRANSMEM 81 100
FT TRANSMEM 124 144
FT DOMAIN 145 211 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 212 230
FT TRANSMEM 239 256
FT TRANSMEM 292 309
FT TRANSMEM 321 342
FT TRANSMEM 375 394
FT TRANSMEM 422 440
FT TRANSMEM 457 477
FT TRANSMEM 498 517
FT TRANSMEM 536 554
FT TRANSMEM 555 599
FT DOMAIN 176 176 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 181 181 POTENTIAL.
FT CARBOHYD 184 184 POTENTIAL.
FT CARBOHYD 184 184 POTENTIAL.
SQ SEQUENCE 599 AA; 67001 MW; F60BC913 CRC32;

Query Match 47.6%; Score 59; DB 1; Length 599;
Best Local Similarity 54.5%; Pred. No. 1.19e+00;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 168 WNTDRCSFNSYS 178
QY 481 WDOGRCSISYS 491
I: | | | | |

RESULT 11
ID NTGL_HUMAN STANDARD; PRT; 599 AA.
AC P30531;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SODIUM- AND CHLORIDE-DEPENDENT GABA TRANSPORTER 1.
GN SLC6A1 OR GABATR OR GAT1 OR GAT1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 90353567.
RX NELSON H., MANDIYAN S., NELSON N.;
RT "Cloning of the human brain GABA transporter.";
RL FEBS LETT. 269:181-184(1990).

```

CC -!- FUNCTION: TERMINATES THE ACTION OF GABA BY ITS HIGH AFFINITY
CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- THIS PROTEIN IS THE TARGET OF PSYCHOMOTOR STIMULANTS SUCH AS
CC AMPHETAMINES OR COCAINE.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54673; G31658; -
CC PIR; S11073; S11073.
CC MIM; 137165; -
CC PROSITE; PS00610; NA_NEUROTRAN_SYMPT_1; 1.
CC PROSITE; PS00754; NA_NEUROTRAN_SYMPT_2; 1.
CC PFAM; PF00209; SNF; 1.
CC KW NEUROTRANSMITTER TRANSPORT; TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN;
CC SYMPORT; MULTIGENE FAMILY.
CC DOMAIN 1 52 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 53 73 1 (POTENTIAL).
CC FT TRANSMEM 81 100 2 (POTENTIAL).
CC FT TRANSMEM 124 144 3 (POTENTIAL).
CC FT DOMAIN 145 211 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 212 230 4 (POTENTIAL).
CC FT TRANSMEM 239 256 5 (POTENTIAL).
CC FT TRANSMEM 292 309 6 (POTENTIAL).
CC FT TRANSMEM 321 342 7 (POTENTIAL).
CC FT TRANSMEM 375 394 8 (POTENTIAL).
CC FT TRANSMEM 422 440 9 (POTENTIAL).
CC FT TRANSMEM 457 477 10 (POTENTIAL).
CC FT TRANSMEM 498 517 11 (POTENTIAL).
CC FT TRANSMEM 536 554 12 (POTENTIAL).
CC FT DOMAIN 555 599 CYTOPLASMIC (POTENTIAL).
CC FT CARBOHYD 176 176 POTENTIAL.
CC FT CARBOHYD 181 181 POTENTIAL.
CC FT CARBOHYD 184 184 POTENTIAL.
CC SEQUENCE 599 AA; 67014 MW; BF09D0D6 CRC32;
Query Match 47.6%; Score 59; DB 1; Length 599;
Best Local Similarity 54.5%; Pred. No. 1.19e+00;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Db 168 WNTDRCFSNYS 178
Qy 481 WDQGRCSISYIS 491
RESULT 12
ID Y235_METTA STANDARD; PRT; 630 AA.
AC Q60291;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN MJEC135.
GN MJEC135.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
OC METHANOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBEK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,

RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HORST M.A., KAINE B.P., BORODOVSKI M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESSE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL SCIENCE 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L77118; G1522669; -
CC TIGR; MJEC135; -
CC KW HYPOTHETICAL PROTEIN.
CC SEQUENCE 530 AA; 72441 MW; EAE3B7FE CRC32;
Query Match 47.6%; Score 59; DB 1; Length 630;
Best Local Similarity 63.6%; Pred. No. 1.19e+00;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 588 GWDGSGRSIERY 598
Qy 480 GWDGRCISYI 490
RESULT 13
ID MAT3_HUMAN STANDARD; PRT; 419 AA.
AC P43243;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE MATRIN 3 (FRAGMENT).
GN MATR3.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91236771.
RA BELGRADER P., DEY R., BEREZNEY R.;
RT "Molecular cloning of matrin 3, A 125-kilodalton protein of the
RT nuclear matrix contains an extensive acidic domain.";
RL J. BIOL. CHEM. 266:9893-9899(1991).
CC -!- FUNCTION: MAY PLAY A ROLE IN TRANSCRIPTION OR MAY INTERACT WITH
CC OTHER NUCLEAR MATRIX PROTEINS TO FORM THE INTERNAL FIBROGRANULAR
CC NETWORK.
CC -!- SUBCELLULAR LOCATION: NUCLEAR MATRIX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63483; -; NOT_ANNOTATED_CDS.
CC MIM; 164015; -
CC DR PFAM; PF00076; rtm; 1.
CC KW NUCLEAR PROTEIN.
FT NON_TER 1 1
FT DOMAIN 282 290 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 419 AA; 47035 MW; 73C7FE0B CRC32;
Query Match 46.8%; Score 58; DB 1; Length 419;
Best Local Similarity 50.0%; Pred. No. 1.87e+00;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Db 130 WFOGRCVKVDLSEK 143

```
QY 481 WDQGRCSISYSSR 494
|||||: |||:
RESULT 14
ID MAT3_RAT STANDARD; PRT; 845 AA.
AC P43244;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE MATRIN 3.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RX MEDLINE: 91236771.
RA BELGRADER P., DEY R., BEREZNEY R.;
RT "Molecular cloning of matrin 3. A 125-kilodalton protein of the
RT nuclear matrix contains an extensive acidic domain.";
RL J. BIOL. CHEM. 266:9893-9899(1991).
CC -!- FUNCTION: MAY PLAY A ROLE IN TRANSCRIPTION OR MAY INTERACT WITH
CC OTHER NUCLEAR MATRIX PROTEINS TO FORM THE INTERNAL FIBROGRANULAR
CC NETWORK.
CC -!- SUBCELLULAR LOCATION: NUCLEAR MATRIX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X67687; G5227; -
DR PIR: S25516; S25516.
DR PIR: JN0804; JN0804.
DR PROSITE: PS00191; CYTOCHROME_B5; 1.
DR PROSITE: PS00559; MOLYBDOPTERIN_EUK; 1.
DR PFAM: PF00173; heme_1; 1.
DR PFAM: PF00174; oxidored_molyb; 1.
DR PFAM: PF00175; oxidored_fad; 1.
DR PFAM: PF00970; Cyt_reductase; 1.
DR HSP: Pf7571; 2CND.
DR KW OXIDOREDUCTASE; FLAVOPROTEIN; FAD; NADP; HEME; MOLYBDENUM;
KW NITRATE ASSIMILATION.
FT METAL 116 116 MOLYBDENUM-PTERIN (POTENTIAL).
FT DOMAIN 518 591 HEME-BINDING (BY SIMILARITY).
FT BINDING 551 551 HEME LIGAND (BY SIMILARITY).
FT BINDING 575 575 HEME LIGAND (BY SIMILARITY).
FT DOMAIN 614 908 FLAVIN-BINDING DOMAIN (BY SIMILARITY).
FT NP_BIND 877 886 NADP (BY SIMILARITY).
SQ SEQUENCE 908 AA; 101525 MW; 7194CDF CRC32;
Query Match 46.88; Score 58; DB 1; Length 908;
Best Local Similarity 45.5%; Pred. No. 1.87e+00;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Db 898 GWDLQRCVVEF 908
QY 480 GWDQGRCSISY 490
|||||: |||:
Search completed: Thu Jul 8 18:59:24 1999
Job time : 8 secs.

QY 558 WFOGRCVKVDLSEK 571
|||||: |||:
RESULT 15
ID NIA_USTWA STANDARD; PRT; 908 AA.
AC Q05531.
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE NITRATE REDUCTASE (NADPH) (EC 1.6.6.3) (NR).
GN NAR1.
OS USTILAGO MAYDIS (SMUT FUNGUS).
OC EUKARYOTA; FUNGI; BASIDIOMYCOTA; USTILAGINOMYCETES;
OC USTILAGINOMICETIDAE; USTILAGINACEAE; USTILAGO.
RN [1]
RX MEDLINE: 93380675.
RA BANKS G.R., HOLDEN D.W., KANUGA N., SHELTON P., SPANOS A.;
RT "The ustilago maydis nar1 gene encoding nitrate reductase activity:
RT sequence and transcriptional regulation.";
RL GENE 131:69-78(1993).
CC -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -!- CATALYTIC ACTIVITY: NADPH + NITRATE -> NADP(+) + NITRITE + H(2)O.
CC -!- COFACTOR: EACH SUBUNIT CONTAINS 1 EQUIVALENT OF FAD, HEME IRON,
CC AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME GROUP IS
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 18:59:42 1999; MasPar time 7.32 Seconds

Tabular output not generated. 119.371 Million cell updates/sec

Title: >US-09-041-236-2

Description: (480-495) from US09041236.pep (26 of 45)

Perfect Score: 124

Sequence: 1 GWDQGRCSIYSSRS 16

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 26.896; Variance 33.082; scale 0.813

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	124	100.0	666	4	SEMAPHORIN L	7.01e-17
2	66	53.2	505	4	PUTATIVE GABA-GATED CH	5.21e-02
3	66	53.2	506	4	GABA RECEPTOR EPSILON	5.21e-02
4	66	53.2	506	4	GABA-A RECEPTOR EPSILON	5.21e-02
5	62	50.0	219	4	DJ774124.1 (SIMILAR TO 3.79e-01	3.79e-01
6	61	49.2	384	11	GABA-A RECEPTOR EPSILON	6.16e-01
7	61	49.2	801	5	INTEGRIN PRECURSOR	6.16e-01
8	61	49.2	2146	3	POLYKETIDE SYNTHASE	6.16e-01
9	61	49.2	2146	3	POLYKETIDE SYNTHASE	6.16e-01
10	61	49.2	2561	5	T12D8.1 PROTEIN	6.16e-01
11	60	48.4	228	6	GABA TRANSPORTER GAT-1	9.94e-01
12	60	48.4	344	2	HYGROMYCIN B PHOSPHOTR	9.94e-01
13	60	48.4	506	4	GABRE	9.94e-01
14	60	48.4	598	13	GAT-1 GABA TRANSPORTER	9.94e-01
15	60	48.4	720	13	CDC47P	9.94e-01
16	59	47.6	199	2	ORF4	1.60e+00
17	59	47.6	199	2	ORF4	1.60e+00
18	59	47.6	315	2	HYPOTHETICAL 34.0 KD P	1.60e+00
19	59	47.6	690	5	KATANIN P80 SUBUNIT	1.60e+00
20	58	46.8	217	2	HYPOTHETICAL 25.6 KD P	2.55e+00

21	58	46.8	449	4	043934	ET PUTATIVE TRANSLATIO	2.55e+00
22	58	46.8	553	3	014091	HYPOTHETICAL 61.7 KD P	2.55e+00
23	58	46.8	845	11	035833	MATRIN 3.	2.55e+00
24	58	46.8	2352	5	061240	HRNOTCH PROTEIN.	2.55e+00
25	58	46.8	3097	5	015943	DN-CADHERIN.	2.55e+00
26	57	46.0	288	2	005145	2-HYDROXY-6-KETONONA-2	4.05e+00
27	57	46.0	310	14	098528	SIMILAR TO PBCV-1 ORF	4.05e+00
28	57	46.0	506	5	046087	1-EVIDENCE=PREDICTED B	4.05e+00
29	57	46.0	653	4	060620	KATANIN P80 SUBUNIT.	4.05e+00
30	57	46.0	1264	5	018291	C29A12.4 PROTEIN.	4.05e+00
31	57	46.0	1358	10	023887	ALDEHYDE OXIDASE (EC 1	4.05e+00
32	56	45.2	133	8	032960	ORF133.	6.39e+00
33	56	45.2	262	2	005826	HYPOTHETICAL 27.7 KD P	6.39e+00
34	56	45.2	310	14	098540	GENOME, PARTIAL SEQUEN	6.39e+00
35	56	45.2	504	5	027403	GCM PROTEIN.	6.39e+00
36	56	45.2	658	10	065214	CYSTINE PROTEASE.	6.39e+00
37	56	45.2	1537	13	092072	MODIFICATION METHYLASE	6.39e+00
38	56	45.2	1806	5	045657	K04B12.1 PROTEIN.	6.39e+00
39	55	44.4	104	14	011430	HYPOTHETICAL 12.0 KD P	1.00e+01
40	55	44.4	254	2	051256	GLYCEROL UPTAKE FACILI	1.00e+01
41	55	44.4	430	2	086330	HYPOTHETICAL 47.9 KD P	1.00e+01
42	55	44.4	471	2	006959	REBL PROTEIN.	1.00e+01
43	55	44.4	551	3	008109	CHROMOSOME XV READING	1.00e+01
44	55	44.4	783	5	P92163	INTEGRIN BETA G SUBUNI	1.00e+01
45	55	44.4	957	4	014651	MUCIN (INTESTINAL MUCI	1.00e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	566 AA.
ID	075326			
AC	075326;			
DT	01-NOV-1998 (TREMREL. 08, CREATED)			
DT	01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L.			
GN	SEMAL.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC	CATARRHINI; HOMINIDAE; HOMO.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 98389619.			
RA	LANGHE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;			
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA			
RT	viruses."			
RL	GENOMICS 51:340-350(1998).			
DR	EMBL: AF030698; G3523115; ..			
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;			

Query Match 100.0%; Score 124; DB 4; Length 666;

Best Local Similarity 100.0%; Pred. No. 7.01e-17;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 512 GWDQGRCSIYSSRS 527

QY 480 GWDQGRCSIYSSRS 495

RESULT 2

ID Q99520

AC Q99520;

DT 01-MAY-1997 (TREMREL. 03, CREATED)

DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)

DE PUTATIVE GABA-GATED CHLORIDE CHANNEL.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;

OC CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 97238072.

RA GARRET M., BASCLES L., BOUE-GRABOT E., SARTOR P., CHARRON G.,

RA BLOCH B., MARGOLSKEE R.F.;
RT "An mRNA encoding a putative GABA-gated chloride channel is expressed
in the human cardiac conduction system.";
RL J. NEUROCHEM. 68:1382-1389(1997).
DR EMBL: Y07637; E274573;
DR PFAM: PF00065; neur_chan. 2.
SQ SEQUENCE 505 AA; 57856 MW; 1E6B104C CRC32;

Query Match 53.2%; Score 66; DB 4; Length 505;
Best Local Similarity 63.6%; Pred. No. 5.21e-02;
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 467 WOQGRLCIHVY 477
I:||||| :|
Qy 481 WDQGR-CISYI 490

RESULT 3
ID O15346 PRELIMINARY; PRT; 506 AA.
AC O15346;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GABA RECEPTOR EPSILON SUBUNIT.
GN GABRE.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RA SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE: 97480709.
RA WILKE K., GAUL R., KLAUCK S.M., POUSTKA A.;
RT "A gene in human chromosome band Xq28 (GABRE) defines a putative new
subunit class of the GABA_A neurotransmitter receptor.";
RL GENOMICS 45:1-10(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA HANNA M.C., HALES T.G., KIRKNESS E.F.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: Y09765; E1168937;
DR EMBL: U92283; G2735349;
DR EMBL: U92281; G2735349; JOINED.
DR EMBL: U92282; G2735349; JOINED.
DR PFAM: PF00065; neur_chan. 2.
SQ SEQUENCE 506 AA; 57971 MW; 95B5B1C4 CRC32;

Query Match 53.2%; Score 66; DB 4; Length 506;
Best Local Similarity 63.6%; Pred. No. 5.21e-02;
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 468 WOQGRLCIHVY 478
I:||||| :|
Qy 481 WDQGR-CISYI 490

RESULT 4
ID P78334 PRELIMINARY; PRT; 506 AA.
AC P78334;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GABA-A RECEPTOR EPSILON SUBUNIT.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RA SEQUENCE FROM N.A.
RX MEDLINE: 97192095.
RA DAVIES P.A., HANNA M.C., HALES T.G., KIRKNESS E.F.;
RT "Insensitivity to anesthetic agents conferred by a class of GABA(A)
receptor subunit.";
RL NATURE 385:820-823(1997).

DR EMBL: U66661; G1857126;
DR PFAM: PF00065; neur_chan. 2.
SQ SEQUENCE 506 AA; 57955 MW; 6BC8394B CRC32;

Query Match 53.2%; Score 66; DB 4; Length 506;
Best Local Similarity 63.6%; Pred. No. 5.21e-02;
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 468 WOQGRLCIHVY 478
I:||||| :|
Qy 481 WDQGR-CISYI 490

RESULT 5
ID O75997 PRELIMINARY; PRT; 219 AA.
AC O75997;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-JAN-1999 (TREMBLREL. 09, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE DJ774124.1 (SIMILAR TO PREGNANCY-ASSOCIATED PLASMA PROTEIN A
DE PRECURSOR) (FRAGMENT).
GN DJ774124.1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RA SEQUENCE FROM N.A.
RA GRAHAM D.;
RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AL031290; E1355032;
FT NON_TER 1
SQ SEQUENCE 219 AA; 24366 MW; FOA8FEAB CRC32;

Query Match 50.0%; Score 62; DB 4; Length 219;
Best Local Similarity 55.6%; Pred. No. 3.79e-01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 13 WEQGSCIPV 21
I:||||| :|
Qy 481 WDQGR-CISYI 489

RESULT 6
ID O55209 PRELIMINARY; PRT; 384 AA.
AC O55209;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE GABA-A RECEPTOR EPSILON SUBUNIT (FRAGMENT).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RA SEQUENCE FROM N.A.
RA HANNA M.C., HALES T.G., KIRKNESS E.F.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U92284; G2735329;
FT NON_TER 1
SQ SEQUENCE 384 AA; 45465 MW; E5D0E1B9 CRC32;

Query Match 49.2%; Score 61; DB 11; Length 384;
Best Local Similarity 50.0%; Pred. No. 6.16e-01;
Matches 6; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Db 345 SWQGRCIHVY 356
I:||||| :|
Qy 480 GWDQGR-CISYI 490

RESULT 7
ID P91774 PRELIMINARY; PRT; 801 AA.
AC P91774;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE INTEGRIN RECURSOR
 OS PACIFASTACUS LENUSCULUS (SIGNAL CRAYFISH).
 OC EUKARYOTA; METAZOA; ARTHROPODA; CRUSTACEA; MALACOSTRACA;
 OC EUMALACOSTRACA; EUCARIDA; DECAPODA; PLECOYEMATA; ASTACOIDEA;
 OC ASTACIDAE; PACIFASTACUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA HOLMBLAD T., THORNOVIST P.O., SODERHALL K., JOHANSSON M.W.;
 RL J. EXP. ZOOL. 277:255-261(1997).
 CC -!- FUNCTION: INTEGRINS ARE A LARGE FAMILY OF CELL SURFACE
 CC GLYCOPROTEINS THAT MEDIATE CELL TO CELL & CELL TO MATRIX
 CC ADHESION.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- SIMILARITY: WITH OTHER BETA CHAINS FROM THE INTEGRIN FAMILY OF
 CC CELL-SURFACE RECEPTOR.
 CC EMBL; X98852; E254398; -.
 DR PROSITE; PS00243; INTEGRIN_BETA; 3.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PFAM; PF00008; EGF; 1.
 DR PFAM; PF00362; Integrin_B; 1.
 DR SIGNAL; INTEGRIN; CELL ADHESION; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT.
 FT SIGNAL 1 25 POTENTIAL.
 SQ SEQUENCE 801 AA; 88479 MW; 5FAE39FA CRC32;

Query Match 49.2%; Score 61; DB 5; Length 801;
 Best Local Similarity 46.7%; Pred. No. 6.16e-01;
 Matches 7; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

Db 267 GWDCQARRVIFSTD 281
 Qy 480 GW-DGRCISYSS 493

RESULT 8
 ID O60026 PRELIMINARY; PRT: 2146 AA.
 AC O60026;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE POLYKETIDE SYNTHASE.
 GN PKSP.
 OS ASPERGILLUS FUMIGATUS (SARTORIA FUMIGATA).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES; EUROTIALES;
 OC TRICHOCOMACEAE; MITOSPORIC TRICHOCOMACEAE; ASPERGILLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC45645;
 RA LANGFELDER K., GEHRINGER H., SCHMIDT A., WANNER G., BRAKHAGE A.A.;
 RL SUBMITTED (MAY-1998), TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; Y17317; E1294271; -.
 DR PROSITE; PS00606; B_KETOACYL-SYNTHASE; 1.
 KW TRANSFERASE.
 SQ SEQUENCE 2146 AA; 234607 MW; 73A0200E CRC32;

Query Match 49.2%; Score 61; DB 3; Length 2146;
 Best Local Similarity 42.9%; Pred. No. 6.16e-01;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 1538 GWDSMRCLKKFSGD 1551
 Qy 480 GWDCQRCISYSS 493

RESULT 9
 ID O59897 PRELIMINARY; PRT: 2146 AA.
 AC O59897;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DE POLYKETIDE SYNTHASE.
 GN ALB1.

OS ASPERGILLUS FUMIGATUS (SARTORIA FUMIGATA).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES; EUROTIALES;
 OC TRICHOCOMACEAE; MITOSPORIC TRICHOCOMACEAE; ASPERGILLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B-5233;
 RA TSAI H.-F., CHANG Y.-C., WASHBURN R.G., WHEELER M.H., KWON-CHUNG K.J.;
 RL SUBMITTED (SEP-1997), TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF025541; G3136092; -.
 DR PROSITE; PS00606; B_KETOACYL-SYNTHASE; 1.
 KW TRANSFERASE.
 SQ SEQUENCE 2146 AA; 234553 MW; 972438E5 CRC32;

Query Match 49.2%; Score 61; DB 3; Length 2146;
 Best Local Similarity 42.9%; Pred. No. 6.16e-01;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 1538 GWDSMRCLKKFSGD 1551
 Qy 480 GWDCQRCISYSS 493

RESULT 10
 ID O46025 PRELIMINARY; PRT: 2561 AA.
 AC O46025; O62379;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE T12D8.1 PROTEIN.
 GN T12D8.1.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STEWARD C.;
 RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL NATURE 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MCMURRAY A.;
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; 282094; E1351242; -.
 DR EMBL; 281120; E1351242; JOINED.
 DR EMBL; 281120; E1349319; -.
 DR EMBL; 282094; E1349319; JOINED.
 SQ SEQUENCE 2561 AA; 286099 MW; 8E1BA84D CRC32;

Query Match 49.2%; Score 61; DB 5; Length 2561;
 Best Local Similarity 35.7%; Pred. No. 6.16e-01;
 Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 663 WOHGACEGLYTDQ 676
 Qy 481 WDCQRCISYSS 494

RESULT 11

DC 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GABRE.
GN GABRE.

"Licensing of DNA replication by a multi-protein complex of MCM/P1

RT proteins in Xenopus eggs."
RL EMBO J. 16:3320-3331(1997).
DR EMBL: U44051; G223117;
DR PROSITE: PS00847; MCM_1; 1.
DR PFM; PF00493; MCM; 1.
SQ SEQUENCE 720 AA; 82021 MW; 000495E4 CRC32;

Query Match 48.4%; Score 60; DB 13; Length 720;
Best Local Similarity 50.0%; Pred. No. 9.94e-01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 246 WNIPRCMSVY 255

I: I: I: I:

Qy 481 WDGRCISIIY 490

Search completed: Thu Jul 8 18:59:56 1999
Job time : 14 secs.

THIS PAGE BLANK (USPTO)

WPSREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPorch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:02:57 1999; MasPar time 10.40 Seconds
Tabular output not generated. 16.352 Million cell updates/sec

Title: >US-09-041-236-2
Description: (531-538) from US09041236.pep (27 of 45)
Perfect Score: 66
Sequence: 1 YLSPMES 8

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseg35
1:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 16.172; Variance 44.065; scale 0.367

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	80.3	372	39	W89449 A gida2 polypeptide f	1.71e+01
2	53	80.3	435	39	W89448 A gida2 polypeptide s	1.71e+01
3	47	71.2	113	20	R96573 Hepatitis C virus typ	8.87e+01
4	47	71.2	1014	35	W68474 HIV-1 strain YBF30 po	8.87e+01
5	46	69.7	268	23	W10013 Interleukin-1 beta.	1.16e+02
6	46	69.7	520	22	W15747 Progesterone-regulated g	1.16e+02
7	45	68.2	267	12	R63137 Porcine pro-interleuk	1.51e+02
8	45	68.2	268	1	R96018 Rat IL-1beta protein.	1.51e+02
9	45	68.2	268	8	R42837 Rat IL-1beta.	1.51e+02
10	45	68.2	269	7	R36531 Mouse IL-1.	1.51e+02
11	45	68.2	419	2	R06428 SPI-like protein enco	1.51e+02
12	45	68.2	426	3	R12458 Pregnancy-specific be	1.51e+02
13	44	66.7	205	35	W70222 Leishmania antigen MA	1.97e+02
14	44	66.7	267	18	R91362 Proenkephalin.	1.97e+02
15	44	66.7	462	1	P82080 Active domain of hydr	1.97e+02
16	44	66.7	462	8	R39635 Human 3-hydroxy-3-met	1.97e+02

17	44	66.7	510	18	R88360	Caenorhabditis elegans	1.97e+02
18	44	66.7	555	4	R22925	Truncated hamster HMG	1.97e+02
19	44	66.7	566	37	W81816	Human ZPB protein.	1.97e+02
20	44	66.7	566	10	R55207	Human zona pellucida	1.97e+02
21	44	66.7	887	11	R58609	Hamster HMG-CoA reduc	1.97e+02
22	44	66.7	914	23	W18030	Huntingtin interactin	1.97e+02
23	43	65.2	98	34	W70952	Light chain of monocl	2.56e+02
24	43	65.2	456	20	W05246	Drosophila glutamate	2.56e+02
25	43	65.2	508	6	R30875	Prod. of the insert i	2.56e+02
26	42	63.6	25	4	R23887	N-terminal sequence o	3.32e+02
27	42	63.6	99	24	W13816	Carboxyl-terminal cel	3.32e+02
28	42	63.6	111	6	R28068	Sequence encoded by t	3.32e+02
29	42	63.6	113	20	R96564	Hepatitis C virus typ	3.32e+02
30	42	63.6	145	1	P94865	Der p II antigen from	3.32e+02
31	42	63.6	216	24	W25110	Human cholesterol 7-a	3.32e+02
32	42	63.6	286	18	R88917	Human cholesterol 7-a	3.32e+02
33	42	63.6	309	39	W73639	Mouse B7-2 antigen.	3.32e+02
34	42	63.6	309	13	R67988	Murine B lymphocyte a	3.32e+02
35	42	63.6	309	20	W08468	Mouse B lymphocyte an	3.32e+02
36	42	63.6	313	26	W34452	Rat CD86.	3.32e+02
37	42	63.6	314	16	R82899	Mouse B7-2 alternativ	3.32e+02
38	42	63.6	382	38	W81569	Human lactosyl cerami	3.32e+02
39	42	63.6	482	14	R72601	Truncated human chole	3.32e+02
40	42	63.6	490	39	W89764	Staphylococcus aureus	3.32e+02
41	42	63.6	504	14	R72507	Human cholesterol 7 a	3.32e+02
42	42	63.6	504	14	R73945	Human cholesterol 7 a	3.32e+02
43	42	63.6	504	5	R24482	Cholesterol 7-alpha-h	3.32e+02
44	41	62.1	387	5	R25335	D4 dopamine receptor.	4.29e+02
45	41	62.1	467	9	R48950	Sequence encoded by a	4.29e+02

ALIGNMENTS

RESULT 1

ID W89449 standard; Protein; 372 AA.
AC W89449;
DT 18-WAR-1999 (first entry)
DE A gida2 polypeptide fragment.
KW gida2; Staphylococcus aureus WCUH29; bacterial infection;
KW Helicobacter pylori infection; cancer; ulcer; gastritis; antibacterial;
OS wound treatment; bacterial adhesion; matrix protein.
KW Staphylococcus aureus.
PN EP-889131-A2.
PD 07-JAN-1999.
PE 30-JUN-1998; 305203.
PR 01-JUL-1997; US-051380.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI (SMIK) SMITHKLINE BEECHAM PLC.
PI Burnham M, Debouck CW, Kallender H, Lenox AL, Mooney JL,
PI Palmer LM, Zhong YY;
PI WPI; 99-062662/06.
DR N-PSDB; V82085.
PT New isolated gida2 polypeptide from Staphylococcus aureus - used to
PT diagnose, treat and prevent bacterial infections e.g. S. aureus and
PT H. pylori, related cancers, ulcers and gastritis and to prevent
PT adhesion of bacteria to matrix proteins
PS Claim 1; Page 6; 41pp; English.
CC The present sequence represents a gida2 protein fragment of
CC Staphylococcus aureus WCUH29. The gida2 proteins and nucleic acids
CC are used to treat conditions requiring increased activity or expression
CC of gida2, while conditions (e.g. bacterial infections) requiring
CC inhibition of such activity or expression are treated by administering
CC an antagonist, inhibitory nucleic acid or competitive polypeptide.
CC Infection by S. aureus is treated, but also Helicobacter pylori
CC infections and related cancers, ulcers and gastritis. These antibacterial
CC agents may also be used to treat in-dwelling devices to prevent infection
CC or generally as wound treatments to prevent adhesion of bacteria to
CC matrix proteins.
SQ Sequence 372 AA;

Query Match 80.3%; Score 53; DB 39; Length 372;
Best Local Similarity 83.3%; Pred. No. 1.71e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Db 188 ylncpm 193
   11:111
QY 531 YLSCPM 536

RESULT 2
ID W89448 standard; Protein: 435 AA.
AC W89448;
DT 18-MAR-1999 (first entry)
DE A g1dA2 polypeptide sequence.
KW g1dA2; Staphylococcus aureus WCuH29; bacterial infection;
KW Helicobacter pylori infection; cancer; ulcer; gastritis; antibacterial;
KW wound treatment; bacterial adhesion; matrix protein.
OS Staphylococcus aureus.
PN EP-889131-A2.
PD 07-JAN-1999. 305203.
PF 30-JUN-1998; US-051380.
PR 01-JUL-1997; US-051380.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
PI Burnham M, Debouck CM, Kallender H, Lenox AL, Mooney JL,
PI Palmer LM, Zhong YY;
DR WPI: 99-062662/06.
DR N-PSDB: V82084.
PT New isolated g1dA2 polypeptide from Staphylococcus aureus - used to
PT diagnose, treat and prevent bacterial infections e.g. S. aureus and
PT H. pylori, related cancers, ulcers and gastritis and to prevent
PT adhesion of bacteria to matrix proteins
PS Claim 1; Page 4-5; 41pp; English.
CC The present sequence represents a g1dA2 polypeptide of Staphylococcus
CC aureus WCuH29. The g1dA2 proteins and nucleic acids are used to treat
CC conditions requiring increased activity or expression of g1dA2, while
CC conditions (e.g. bacterial infections) requiring inhibition of such
CC activity or expression are treated by administering an antagonist,
CC inhibitory nucleic acid or competitive polypeptide. Infection by
CC S. aureus is treated, but also Helicobacter pylori infections and
CC related cancers, ulcers and gastritis. These antibacterial agents may
CC also be used to treat in-dwelling devices to prevent infection or
CC generally as wound treatments to prevent adhesion of bacteria to matrix
CC proteins.
CC Sequence 435 AA:
Query Match 80.3%; Score 53; DB 39; Length 435;
Best Local Similarity 83.3%; Pred. No. 1.71e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 188 ylncpm 193
   11:111
QY 531 YLSCPM 536

RESULT 3
ID R96573 standard; peptide; 113 AA.
AC R96573;
DT 14-MAR-1997 (first entry)
DE Hepatitis C virus types 9a(7a) isolates FR1 amino acids 2645-2757.
KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
KW PCR; primer; probe; antibody; infection.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT misc_difference 12
FT /label.. Lev, Pro, Gln, Arg
FN W09613590-A2.
PD 09-MAY-1996.
PF 23-OCT-1995; E04155.
PR 21-OCT-1994; EP-870166.
PR 28-JUN-1995; EP-870076.
PA (INNO-) INNOGENETICS NV.
PI Maertens G, Stuyver L;
DR WPI: 96-251460/25.
DR N-PSDB: T27984.
PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:stype

- used to develop probes and primers for new sub:types and vaccines
to prevent and treat infection
Claim 25; Fig 3: 150pp; English.
The sequences R96526-R96578 represent novel sequences isolated from
hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5',
untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
This sequence represents amino acids 2645-2757 from the HCV types 9a and
7a isolates FR1.
The new HCV types were isolated from patients with chronic HCV from the
Benelux countries, France, Cameroon and Vietnam, because of their
aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
sequenced either directly or partially and used to classify the new
viruses into (sub)types based on comparison with known sequences.
The sequences were used to generate the peptides R96424-R96524. The
sequences can also be used to synthesise probes and primers for the
detection of HCV in a sample. The polypeptides can be used to detect
anti-HCV antibodies, for HCV typing or to prevent HCV infections.
Sequence 113 AA;
Query Match 71.2%; Score 47; DB 20; Length 113;
Best Local Similarity 50.0%; Pred. No. 8.87e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 15 ylsqldp 22
   11111111
QY 531 YLSCPMES 538

RESULT 4
ID W68474 standard; Protein; 1014 AA.
AC W68474;
DT 08-DEC-1998 (first entry)
DE HIV-1 strain YBF30 pol protein.
KW HIV-1 strain YBF30; antibody; oligonucleotide; diagnosis; immunisation;
KW infection; typing; pol.
OS Human immunodeficiency virus type 1.
PN FR2756843-A1.
PD 12-JUN-1998.
PF 09-DEC-1996; 015087.
PR 09-DEC-1996; PR-015087.
PA (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA (INSP ) INST PASTEUR.
PI Barre-Sinoussi F, Lousserat-Ajaka I, Mauciere P, Saragosti S,
PI Simon F;
DR WPI: 98-336114/30.
DR N-PSDB: V60754.
PT Non-M, non-O HIV-1 strain YBF30 - useful for diagnosis and
PT immunisation
PS Claim 3; Page 29-33; 85pp; French.
CC This sequence represents the pol protein from the non-M (major), non-O
CC (Outlier) HIV-1 strain YBF30 (CNCM I-1753), isolated from the Cameroon.
CC The HIV strain (see V60751 for complete genome), peptides, antibodies and
CC oligonucleotides derived from it (see V60752-V60798 and W68473-W68482)
CC are used for diagnosis of or immunisation against non-M, non-O HIV-1
CC infections. The oligonucleotides, peptides and antibodies can also be
CC used for typing HIV strains.
CC Sequence 1014 AA;
Query Match 71.2%; Score 47; DB 35; Length 1014;
Best Local Similarity 57.1%; Pred. No. 8.87e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 277 yfscpld 283
   11111111
QY 531 YLSCPME 537

RESULT 5
ID W10013 standard; Protein; 268 AA.

```


AC W10013;
 DT 15-SEP-1997 (first entry)
 DE Interleukin-1 beta
 KW horse; antibody; alpha; beta; IL-1; research; reagent; inflammation;
 KW peptide; antigen.
 OS Equus caballus.
 PN J09131191-A.
 PD 20-MAY-1997.
 PF 01-AUG-1996; 203551.
 PR 04-SEP-1995; JP-226133.
 RA (HITB) HITACHI CHEM CO LTD.
 DR WPI; 97-327277/30.
 DR N-PSDB; T70089.
 PT Horse interleukin 1 peptide and related DNA - useful for preparation
 PT of anti-horse interleukin 1 antibody used in studying horse
 PT inflammatory diseases
 PS Claim 1; Page 12-13; 14pp; Japanese.
 CC This sequence is a horse interleukin-1 (IL-1) beta chain. Horse IL-1
 CC peptides consisting of at least 5 contiguous amino acids from the alpha
 CC or beta chain are claimed and can be used to prepare an anti-horse IL-1
 CC antibody. The peptide and antibodies are useful as research reagents for
 CC studying the relationship between horse inflammatory diseases and horse
 CC IL-1.
 CC Sequence 268 AA;
 SQ Sequence 268 AA;
 Query Match 69.7%; Score 46; DB 23; Length 268;
 Best Local Similarity 83.3%; Pred. No. 1.16e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 183 ylsccm 188
 Qy 531 YLSCPM 536
 RESULT 6
 ID W15747 standard; Protein; 520 AA.
 AC W15747;
 DT 11-JUN-1997 (first entry)
 DE Progesterin-regulated gene-1 (PRG-1) polypeptide.
 KW Progesterin-regulated gene 1; PRG1; breast cancer; tumour;
 KW hyperproliferation; metabolic disease; diabetes; cell cycle;
 KW antiproliferative; gene therapy.
 OS Homo sapiens.
 PN W09715674-A1.
 PD 01-MAY-1997.
 PF 24-OCT-1996; AU06569.
 PR 24-OCT-1995; AU-006144.
 PR 19-JUL-1996; AU-001128.
 PA (GARV) GARVAN INST MEDICAL RES.
 PI Hamilton JA, Watts CKW;
 DR WPI; 97-259026/23.
 DR N-PSDB; T60548.
 PT Progesterin-regulated gene - indicative of a subject's progesterin
 PT responsiveness, also for treatment of metabolic disease or
 PT hyperproliferation etc
 PS Example; Fig 2B; 50pp; English.
 CC The polypeptide (W15747) encoded by the novel human progesterin-
 CC regulated gene PRG1 (T60548) appears to have a novel enzymatic
 CC activity that may be useful as a readily detectable marker for
 CC progesterin responsiveness. The 520 amino acid protein shows high
 CC similarity to human liver PFK/FBP and has several consensus
 CC phosphorylation sites. Regulating the enzymatic activity of
 CC PRG1 by using (ant)agonists, ribozymes or by gene therapy can
 CC be used to treat hyperproliferation, e.g. cancer and skin
 CC diseases. Anti-PRG1 antibodies can be used to assess a subject's
 CC responsiveness to progesterin; high levels of PRG1 indicate that a
 CC breast tumour will be treatable with progesterin. PRG1 is probably
 CC involved in glycolytic, gluconeogenic and lipogenic pathways, and
 CC may be useful in the treatment of diabetes and other disorders of
 CC energy metabolism, and to control reproductive function.
 CC Sequence 520 AA;
 Query Match 69.7%; Score 46; DB 22; Length 520;

Best Local Similarity 50.0%; Pred. No. 1.16e+02;
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 409 ylkcp1ht 416
 Qy 531 YLSCPMES 538
 RESULT 7
 ID R63137 standard; Protein; 267 AA.
 AC R63137;
 DT 13-JUN-1995 (first entry)
 DE Porcine pro-interleukin-1 beta.
 KW Porcine interleukin-1 beta; IL-1beta; growth factor;
 KW immunostimulant; exon amplification.
 OS Sus scrofa.
 FH Key Location/Qualifiers
 FT region 1..113
 FT /label= pro-region
 FT modified_site 59..61
 FT /note= "potential N-glycosylation site"
 FT protein 114..267
 FT /label= IL-1beta
 PN NL9301929-A.
 PD 17-OCT-1994.
 PF 05-NOV-1993; 001929.
 PR 29-MAR-1993; EP-200907.
 PA (REGA-) STICHTING REGA VZW.
 PI Billiau AJDA, Vandenbroeck K;
 DR WPI; 94-322838/40.
 DR N-PSDB; Q73473.
 DT DNA coding for porcine interleukin 1-beta - and new recombinant
 DT porcine interleukin 1-beta useful as growth factor and
 DT immunostimulant
 PS Claim 1; Fig 3b; 40pp; Dutch.
 CC The genomic DNA sequence coding for porcine interleukin-1 beta was
 CC determined (Q73473). The mature porcine IL-1beta polypeptide is useful
 CC as a metabolic growth factor in healthy pigs, e.g. modifying glucose
 CC turnover, energy consumption and lipid metabolism. It is also useful
 CC as an immunostimulant, esp. as a vaccine adjuvant.
 CC Sequence 267 AA;
 SQ Sequence 267 AA;
 Query Match 68.2%; Score 45; DB 12; Length 267;
 Best Local Similarity 83.3%; Pred. No. 1.51e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 182 ylsccm 187
 Qy 531 YLSCPM 536
 RESULT 8
 ID P96018 standard; Protein; 268 AA.
 AC P96018;
 DT 07-SEP-1990 (first entry)
 DE Rat IL-1beta protein.
 KW Rat; IL-1betaprotein; precursor.
 OS Rattus sp.
 FH Key Location/Qualifiers
 FT protein 1..268
 FT /product=rat IL-1beta precursor.
 FT protein 117..268
 FT /product=rat IL-1beta.
 PN J01071493-A.
 PD 16-MAR-1989.
 PF 11-SEP-1987; 229310.
 PR 11-SEP-1987; JP-229310.
 PA (SAKA) Otsuka Pharm KK.
 DR WPI; 89-125944/17.
 DR N-PSDB; N90914.
 DT Rat IL-1 beta gene - encoding rat IL-1 beta or IL-1 beta precursor.
 PS Disclosure; 2; 14pp; Japanese.
 CC Bacterial lipopolysaccharide was administered to Sprague-Dawley rats, and

CC brain tissue removed 6 hrs later. After treatment the mRNA was isolated
 CC by column chromatography using oligo (dT)-cellulose.
 SQ Sequence 268 AA;

Query Match 68.2%; Score 45; DB 1; Length 268;
 Best Local Similarity 83.3%; Pred. No. 1.51e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 184 ylsbcm 189
 QY 531 YLSCPM 536

RESULT 9

ID R42837 standard; Protein; 268 AA.
 AC R42837; 1994 (first entry)
 DT 09-MAR-1994
 DE Rat IL-1beta.
 KW Interleukin-1; IL-1; detection; monoclonal antibody;
 KW hybridoma.
 OS Rat rattus.
 FH Key
 FT Cds Location/Qualifiers
 FT 66..872
 FT /tag= a
 FT /product= IL-1beta
 FT
 PN J05244990-A.
 PD 24-SEP-1993.
 PF 04-MAR-1992; 047240.
 PR 04-MAR-1992; JP-047240.
 PA (SAKA) OTSUKA PHARM CO LTD.
 DR WPI: 93-338945/43.
 DR N-PSDB: Q49458.
 PT Anti-rat interleukin-1 monoclonal antibody - prepd. from
 PT hybridoma, useful for accurate detection of rat IL-1
 PS Claim 1; Fig 2-3; 21pp; Japanese.
 CC An anti-rat IL-1 monoclonal antibody, produced by a hybridoma formed
 CC by fusion of a mammalian cell immunised by rat IL-1alpha or rat IL-
 CC 1beta, and a mammalian marrow cell, has specific reactivity with rat
 CC IL-1alpha or rat IL-1beta. The antibody may be used in
 CC the detection of rat IL-1alpha or beta.
 SQ Sequence 268 AA;

Query Match 68.2%; Score 45; DB 8; Length 268;
 Best Local Similarity 83.3%; Pred. No. 1.51e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 184 ylsbcm 189
 QY 531 YLSCPM 536

RESULT 10

ID R36531 standard; Protein; 269 AA.
 AC R36531;
 DT 12-AUG-1993 (first entry)
 DE Mouse IL-1.
 KW Interleukin; plamsid; expression; macrophage.
 OS Mus musculus.
 PN J05084084-A.
 PD 06-APR-1993.
 PF 11-JAN-1991; 044540.
 PR 11-JAN-1991; JP-044540.
 PA (HONS) YAKULT HONSHA KK.
 DR WPI: 93-148498/18.
 DR N-PSDB: Q40567.
 PT Mouse interleukin-1-producing plasmid - contains mouse
 PT interleukin-1 cDNA fragment originated from mouse macrophage
 PS Disclosure: Page 13-14; 27pp; Japanese.
 CC A mouse interleukin-1 (mIL-1) producing plasmid contg. a mIL-1 cDNA
 CC fragment from mouse macrophage is new. The mIL-1 producing plasmid
 CC has a high productivity.
 SQ Sequence 269 AA;

Query Match 68.2%; Score 45; DB 7; Length 269;
 Best Local Similarity 83.3%; Pred. No. 1.51e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 185 ylsbcm 190
 QY 531 YLSCPM 536

RESULT 11

ID R06428 standard; protein; 419 AA.
 AC R06428;
 DT 17-DEC-1990 (first entry)
 DE SP1-like protein encoded by clone hps11.
 KW pregnancy-specific protein; carcinoembryonic antigen; CEA;
 KW fertility; abortion; placenta.
 OS Homo sapiens.
 FH Key
 FT peptide Location/Qualifiers
 FT 1..34
 FT /label=signal peptide
 FT domain 144..236
 FT /label=Rln
 FT /note="repeat unit"
 FT domain 237..329
 FT /label=R2n
 FT /note="repeat unit"
 FT region 330..414
 FT /label=R2c
 FT modified_site 61..63
 FT modified_site 104..106
 FT modified_site /label=N-glycos_site
 FT modified_site 111..113
 FT modified_site /label=N-glycos_site
 FT modified_site 199..201
 FT modified_site /label=N-glycos_site
 FT modified_site 259..261
 FT modified_site /label=N-glycos_site
 FT modified_site 268..270
 FT modified_site /label=N-glycos_site
 FT modified_site 303..305
 FT modified_site /label=N-glycos_site
 FT peptide 37..47
 FT peptide /label=tryptic peptide
 FT peptide 74..98
 FT peptide /label=tryptic peptide
 FT peptide 201..208
 FT peptide /label=tryptic peptide
 FT peptide 345..364
 FT peptide /label=tryptic peptide
 FT region 35..50
 FT /label=alpha helix
 FT region 160..170
 FT /label=alpha helix
 FT domain 1..143
 FT /label=glycosylation domain
 FT misc_difference 41
 FT /note="different from that determined by
 FT N-terminal analysis of pure SPI"
 FT misc_difference 43
 FT /note="different from that determined by
 FT N-terminal analysis of pure SPI"
 FT
 PN W09007937-A.
 PD 26-JUL-1990.
 PF 11-JAN-1990; U00285.
 PR 18-JAN-1989; US-298638.
 PR 07-AUG-1989; US-390409.
 PA (OKLA-) OKLAHOMA MED RES FO.
 PI Chan WX; 253860/33.
 PT WPI; 90-253860/33.
 PT N-PSDB; Q05639.
 PT pregnancy specific proteins, genes and antibodies - for use in
 PT diagnosis and in compsns. having immunosuppressive and growth

```
PT promoting activities
PS Claim 10; Fig 2; 78pp; English.
CC hPS11 is a clone encoding a placental pregnancy-specific protein
CC (SP1) also clone as pregnancy-specific beta glycoprotein (PSBG),
CC detected in placenta and in testis. The sequence and Abs specific
CC for it can be used in diagnosis, pregnancy testing and monitoring
CC of tumours. The protein has immunosuppressive activity and growth
CC promoting activity and can be used for enhancing fertility in
CC females, viability of a foetus etc.
CC See also R06429-34.
SQ Sequence 419 AA;

    Query Match      68.2%; Score 45; DB 2; Length 419;
    Best Local Similarity 62.5%; Pred. NO. 1.51e+02;
    Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 351 y1scsads 358
   ||||: :|
QY 531 YLSCPME5 538

RESULT 12
ID R12458 standard; Protein: 426 AA.
AC R12458;
DT 30-AUG-1991 (first entry)
DE Pregnancy-specific beta-glycoprotein 95.
KW PSG; pregnancy-specific beta glycoprotein; PSG95; antibodies;
KW gestational trophoblastic disease; hydatidiform mole; choriocarcinoma.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..34
FT domain /label= leader_peptide
FT domain 1..143
FT domain /label= N_domain
FT domain 144..236
FT domain /label= A1_domain
FT domain 837..329
FT domain /label= A2_domain
FT domain 330
FT modified_site /label= B2_domain
FT modified_site 60..62
FT modified_site /label= glycosylation_site
FT modified_site 104..106
FT modified_site /label= glycosylation_site
FT modified_site 111..113
FT modified_site /label= glycosylation_site
FT modified_site 199..201
FT modified_site /label= glycosylation_site
FT modified_site 259..261
FT modified_site /label= glycosylation_site
FT modified_site 268..270
FT modified_site /label= glycosylation_site
FT modified_site 301..303
FT modified_site /label= glycosylation_site
US7536101-A.
PD 14-MAY-1991.
PF 08-JUN-1990; 127407.
PR 08-JUN-1990; US-536101.
PA (USSH ) NAT INST OF HEALTH.
PI Chou J;
DR WPI: 91-185479/25.
DR P-PSDB; R12458.
DE Pregnancy-specific beta-glyco-protein, DNA, RNA and antibodies -
PT used for detecting gestational trophoblastic disease, e.g.
PT hydatidiform mole or choriocarcinoma
PS Disclosure: Fig 1; 28pp; English.
CC Six invariant cysteine residues (not indicated on sequence in
CC specification) are important in disulfide bond formation.
CC The protein is identical to the protein encoded by PSG1d and FL-NCA,
CC but differ at the C-terminal from PSGC, due to alternative splicing.
CC Antibodies to the protein can be used for detection of gestational
CC trophoblastic diseases such as hydatidiform mole or choriocarcinomas.
CC See also Q12092-94

SQ Sequence 426 AA;
    Query Match      68.2%; Score 45; DB 3; Length 426;
    Best Local Similarity 62.5%; Pred. NO. 1.51e+02;
    Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 351 y1scsads 358
   ||||: :|
QY 531 YLSCPME5 538

RESULT 13
ID W70222 standard; Protein: 206 AA.
AC W70222;
DT 13-NOV-1998 (first entry)
DE Leishmania antigen MAPS-1A protein.
KW Leishmania antigen; immune response; infection detection; therapy;
KW humoral response induction; cellular response induction; cancer;
KW interleukin-12 production.
OS Leishmania tropica.
PN WO9835045-A2.
PD 13-AUG-1998.
PF 12-FEB-1998; U03002.
PR 27-AUG-1997; US-920609.
PR 12-FEB-1997; US-798841.
PA (CORI-) CORIXA CORP.
PI Campos-Neto A, Dillon DC, Reed SG, Skeiky YA, Webb JR;
DR WPI: 98-447242/38.
DR N-PSDB; V47566.
PT New immunogenic fragments of Leishmania antigens and related nucleic
PT acid, vectors and host cells - are useful for diagnosis, prevention
PT and treatment of leishmaniasis, also to induce production of
PT interleukin-12 generally.
PS Claim 13; Page 107-108; 194pp; English.
CC This sequence represents a Leishmania antigen (LAG) of the invention,
CC designated MAPS-1A. Compositions and vaccines containing the protein are
CC used to generate a protective or therapeutic immune response against the
CC Leishmania species donovani, chagasi, infantum, major, amazonensis,
CC braziliensis, panamensis, tropica or guayanensis. They can also be used
CC to detect infection (in a skin test). The compositions induce a humoral
CC and/or cellular response, specifically of Th1 type, particularly
CC including induction of interleukin-12 (IL-12) production. They may thus
CC be used more generally to treat any condition (e.g. bacterial, viral or
CC protozoal infection, or cancer) which responds to IL-12.
SQ Sequence 206 AA;
    Query Match      66.7%; Score 44; DB 35; Length 206;
    Best Local Similarity 57.1%; Pred. NO. 1.97e+02;
    Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 81 lacsmads 87
   |.:|.:|
QY 532 LSCPME5 538

RESULT 14
ID R91362 standard; protein: 267 AA.
AC R91362;
DT 09-OCT-1996 (first entry)
DE Proenkephalin.
KW Proenkephalin; antibody; PE; apoptosis inducer; opiod receptor; cancer;
KW monoclonal; cytoplasmic anchor; nuclear localisation signal; cataract;
KW opiod peptide precursor; OPP; lens cell; cell cycle arrest agent;
KW chemotherapeutic; genotoxic; cell cycle synchronisation; therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 98..110
FT key /note= "Antibody PE-2 (see R91362) recognition sequence"
FT region 155..161
FT key /note= "Recognition sequence for antibodies PE-14, PE-15,
FT PE-16, PE-17, PE-18, PE-19, PE-23 and PE-25 (see
FT R91363 and R91368-R91373)"
FT region 175..185
```

FT /note= "Recognition sequence for antibodies PE-13, PE-20,
FT PE-21, PE-22 and PE-24 (see R91364 and
FT R95283-R95286)"
FT 187..192
FT /note= "PE-1 (see R91365) recognition sequence"
PN WO9606863-A1.
PD 07-MAR-1996.
PF 30-AUG-1995; G02037.
PR 30-AUG-1994; GB-017444.
PR 23-SEP-1994; GB-019285.
PA (UYDU-) UNIV DUNDEE.
PI Boettger A, Dewar DA, Prescott A, Spruce BA;
DR WPI; 96-160311/16.
PT New agents for inducing apoptosis in cells - which are able to
PT modulate a biochemical pathway in which prods. of opioid peptide
PT precursor genes participate.
PS Claim 6; Fig 4; 64pp; English.
CC This sequence represents proenkephalin (PE). Epitopes within this
CC sequence are recognised by anti-PE antibodies (see R91363-R91366,
CC R91368-R19373 and R95283-R95286 for partial antibody sequences). The PE
CC antibodies are used in an agent for inducing apoptosis in cells. The
CC agents preferably neutralise this sequence or its proteolytic
CC derivatives, increase the level of, activate or mimic nuclear PE, and act
CC as an antagonist to receptors related or identical to the delta and kappa
CC opioid receptors. The agent is preferably a neutralising monoclonal
CC antibody (such as those represented by R91368-R19373 and R95283-R95286),
CC or a fragment or allelic form of one of these antibodies. A cytoplasmic
CC anchor, or a nuclear localisation signal may also be included in the
CC agent. The agent is able to modulate a biochemical pathway in a cell in
CC which products of opioid peptide precursor (OPP) genes participate in
CC order to induce the cell to apoptose. The agents can be used for the
CC treatment of cancer or for inducing apoptosis in lens cells following a
CC cataract operation. These agents promote apoptosis of proliferating
CC cells with less, or no, effect on normal mature cell types. The agents
CC may be administered in combination with a genotoxic or cell cycle arrest
CC agent. Alternatively, the agent may be complexed with a
CC chemotherapeutic, irradiation or cell cycle arrest (synchronisation)
CC agent.
SQ Sequence 267 AA;

Query Match 66.7%; Score 44; DB 18; Length 267;
Best Local Similarity 57.1%; Pred. No. 1.97e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 45 flacvme 51
! : ! : ! : ! :
Qy 531 YLSCPME 537

RESULT 15
ID P82080 standard; protein; 462 AA.
AC P82080;
DT 23-OCT-1990 (first entry)
DE Active domain of hydroxymethyl glutaryl coenzyme A reductase
KW hydroxymethyl glutaryl coenzyme A reductase (HMG-CoAR);
KW cholesterol biosynthesis; ds.
OS synthetic.
PN EP-292803-A.
PD 30-NOV-1988.
PF 13-MAY-1988; 107661.
PR 23-MAY-1987; DE-717437.
PA (FARH) Hoeschst AG.
PI Uhlmann E, Hein F;
DR WPI; 88-339392/48.
DR N-PSDB; N82030.
PT Hydroxymethyl glutaryl coenzyme A reductase like protein prodn -
PT by synthesising encoding gene, incorporating into expression
PT plasmid and expressing protein in host organism
PS Example; Page 20; 24pp; German.
CC Sequence encodes the active domain of HMG-CoAR (C-terminal amino
CC acids of the natural enzyme). The encoding DNA is constructed to
CC include many restriction sites. This facilitates its use in
CC structure-activity relationship studies of enzyme and inhibitors

CC through the use of mutant forms of the enzyme.
SQ Sequence 462 AA;
Query Match 66.7%; Score 44; DB 1; Length 462;
Best Local Similarity 62.5%; Pred. No. 1.97e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 366 yisctmps 373
! : ! : ! : ! :
Qy 531 YLSCPME 538

Search completed: Thu Jul 8 19:03:16 1999
Job time : 19 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:02:27 1999; MasPar time 6.17 Seconds

Tabular output not generated. 51.992 Million cell updates/sec

Title: >US-09-041-236-2

Description: (531-538) from US09041236.pep (27 of 45)

Perfect Score: 66

Sequence: 1 YLSCPWES 8

Scoring table: PAM 150

Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: pir60

1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.944; Variance 26.125; scale 0.840

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	54	81.8	284	2	E71613	2.40e-01
2	53	80.3	435	2	A69632	glucose-inhibited div 4.06e-01
3	51	77.3	457	2	S45885	probable membrane pro 1.14e+00
4	50	75.8	932	2	A31898	hydroxymethylglutaryl 1.89e+00
5	49	74.2	337	2	A49970	glycine receptor alph 3.12e+00
6	49	74.2	437	2	S34469	gamma-aminobutyric ac 3.12e+00
7	49	74.2	451	2	B49970	glycine receptor alph 3.12e+00
8	49	74.2	452	2	S12381	glycine receptor alph 3.12e+00
9	49	74.2	452	2	S14816	glycine receptor alph 3.12e+00
10	49	74.2	452	2	S18836	glycine receptor alph 3.12e+00
11	49	74.2	464	2	A23682	glycine receptor alph 3.12e+00
12	48	72.7	413	1	S39554	histidine decarboxyla 5.12e+00
13	48	72.7	502	2	A70988	hypothetical protein 5.12e+00
14	48	72.7	1039	2	S46347	pol polyprotein - sim 5.12e+00
15	47	71.2	98	2	S78727	protein YLL018c-a - y 8.34e+00
16	47	71.2	217	2	S46354	pol polyprotein - sim 8.34e+00
17	47	71.2	445	2	T03578	probable glucose-inhi 8.34e+00
18	47	71.2	449	2	S12382	glycine receptor alph 8.34e+00
19	47	71.2	451	2	C49970	glycine receptor alph 8.34e+00
20	47	71.2	457	2	S20662	glycine receptor alph 8.34e+00
21	47	71.2	515	2	S56784	hypothetical protein 8.34e+00
22	47	71.2	1027	1	GNLJ51	pol polyprotein - sim 8.34e+00
23	46	69.7	70	2	S74243	6-phosphofructo-2-kin 1.35e+01

24 46 69.7 161 2 E71078 hypothetical protein 1.35e+01
 25 46 69.7 214 2 JC5846 interleukin-1 beta - 1.35e+01
 26 46 69.7 238 2 S62582 hypothetical protein 1.35e+01
 27 46 69.7 262 2 S64050 hypothetical protein 1.35e+01
 28 46 69.7 401 2 S76788 histidine--trNA ligas 1.35e+01
 29 46 69.7 440 2 JC2065 6-phosphofructo-2-kin 1.35e+01
 30 46 69.7 469 2 A40800 6-phosphofructo-2-kin 1.35e+01
 31 46 69.7 469 2 JC5871 6-phosphofructo-2-kin 1.35e+01
 32 46 69.7 469 2 S77704 6-phosphofructo-2-kin 1.35e+01
 33 46 69.7 469 2 JC1470 6-phosphofructo-2-kin 1.35e+01
 34 46 69.7 470 2 JC2064 6-phosphofructo-2-kin 1.35e+01
 35 46 69.7 471 2 A44872 6-phosphofructo-2-kin 1.35e+01
 36 46 69.7 471 2 S12732 6-phosphofructo-2-kin 1.35e+01
 37 46 69.7 471 1 KIRTFB 6-phosphofructo-2-kin 1.35e+01
 38 46 69.7 519 2 JC4826 6-phosphofructo-2-kin 1.35e+01
 39 46 69.7 531 2 A31780 transcription factor 1.35e+01
 40 46 69.7 544 2 S15664 transcription factor 1.35e+01
 41 46 69.7 716 2 T01802 hypothetical protein 1.35e+01
 42 46 69.7 907 2 T00947 hypothetical protein 1.35e+01
 43 46 69.7 1585 2 T01060 hypothetical protein 1.35e+01
 44 45 68.2 419 2 A33258 pregnancy-specific gl 2.16e+01
 45 45 68.2 653 2 T03102 semaphorin homolog A3 2.16e+01

ALIGNMENTS

RESULT 1

ENTRY E71613 #type complete
 TITLE hypothetical protein PFB0490c - malaria parasite (Plasmodium falciparum)
 ORGANISM #formal_name Plasmodium falciparum
 DATE 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 13-Nov-1998

ACCESSIONS E71613
 REFERENCE A71600
 #authors Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Shalloo, S.; Mason, T.; Yu, K.; Fujii, C.; Pederson, J.; Shen, K.; Jing, J.; Aston, C.; Lai, Z.; Schwartz, D.C.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.; Fraser, C.M.; Adams, M.D.; Venter, J.C.; Hoffman, S.L.

#journal Science (1998) 282:1126-1132
 ~#title Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 #accession E71613
 #status preliminary; nucleic acid sequence not shown; translation not shown

##molecule_type DNA
 ##residues 1-284 #label GAR
 ##cross-references GB:AE001398; GB:AE001362; NID:G3845197; PID:G3845198; TIGR:PFB0490c

GENETICS
 #experimental_source clone 3D7
 #gene PFB0490c
 #length 284 #molecular-weight 32541 #checksum 7578

Query Match 81.8%; Score 54; DB 2; Length 284;
 Best Local Similarity 71.4%; Pred. No. 2.40e-01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 11 YLSCPMD 17
 |||||:
 QY 531 YLSCPME 537

RESULT 2

ENTRY A59632 #type complete
 TITLE glucose-inhibited division protein gid - Bacillus subtilis
 ORGANISM #formal_name Bacillus subtilis
 DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 30-Oct-1998
 ACCESSIONS A59632; S61497
 REFERENCE A59580

```

#authors      Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
              Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
              Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
              A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
              Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
              Chol, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
              Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
              Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
              Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
              M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
              S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
              Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
              C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
              Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
              Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
              Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
              Kurita, K.; Lapidus, A.; Lardinols, S.; Lauber, J.;
              Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
              Maucel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
              M.; Moestl, D.; Nakai, S.; Nock, M.; Noone, D.; O'Reilly,
              M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro,
              V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
              A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
              Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
              Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.;
              Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
              Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
              B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
              Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
              Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
              Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
              Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
              Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
              K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
              Yoshikawa, H.; Danchin, A.

#journal      Nature (1997) 390:249-256
#title        The complete genome sequence of the Gram-positive bacterium
              Bacillus subtilis.
#cross-references GB:299112; GB:AL009126; NID:g2633902; PID:el185204;
              PID:g2633985
#experimental_source strain 168
#accession    A59632
#status       nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues     1-435 #label KUN
#cross-references GB:299112; GB:AL009126; NID:g2633902; PID:el185204;
              PID:g2633985
#experimental_source strain 168
#accession    S61493
#authors      Slack, F.J.; Serror, P.; Joyce, E.; Sonenshein, A.L.
#journal      Mol. Microbiol. (1995) 15:689-702
#title        A gene required for nutritional repression of the Bacillus
              subtilis dipeptide permease operon.
#cross-references MUID:95302982
#accession    S61497
#molecule_type DNA
#residues     362-435 #label SLA
#cross-references EMBL:U13634; NID:g535347; PID:g1401048
#experimental_source strain JH642
GENETICS
#gene         gld
#length       435 #molecular-weight 48063 #checksum 6908
SUMMARY
Query Match      80.3%; Score 53; DB 2; Length 435;
Best Local Similarity 83.3%; Pred. No. 4,06e-01;
Matches          5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 189 YLNCMP 194
||:||||
Qy 531 YLSCPM 536

RESULT 3
ENTRY   #type complete
TITLE   probable membrane protein YBR029c - yeast (Saccharomyces
              cerevisiae)

ALTERNATE_NAMES cerevisiae)
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change
              06-Feb-1998
ACCESSIONS       S45885; S46559
REFERENCE        S45875
#authors         Grivell, L.A.; de Haan, M.; Maat, M.J.; Smits, P.H.M.
#submission      submitted to the Protein Sequence Database, August 1994
#accession       S45885
#molecule_type DNA
#residues        1-457 #label GRI
#cross-references EMBL:235898; NID:g536239; PID:g536240; MIPS:YBR029c
#experimental_source strain S288C
REFERENCE        S46551
#authors         Smits, P.H.M.; de Haan, M.; Maat, C.; Grivell, L.A.
#journal         Yeast (1994) 10(Suppl.A):S75-S80
#title          The complete sequence of a 33 kb fragment on the right arm of
              chromosome II from Saccharomyces cerevisiae reveals 16 open
              reading frames, including ten new open reading frames, five
              previously identified genes and a homologue of the SCO1
              gene.
#accession       S46559
#status          nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues        1-457 #label SMI
#cross-references EMBL:X76078; NID:g498748; PID:g498757
#experimental_source strain S288C
#note           the nucleotide sequence was submitted to the EMBL Data
              Library, November 1993

GENETICS
#gene          SGD:CDS1
#length        457 #molecular-weight 51822 #checksum 5700
#map_position  2R
#keywords       transmembrane protein
FEATURE
75-94          #domain transmembrane #status predicted #label TW1\
159-179        #domain transmembrane #status predicted #label TW2\
185-209        #domain transmembrane #status predicted #label TW3\
210-228        #domain transmembrane #status predicted #label TW4\
253-271        #domain transmembrane #status predicted #label TW5\
335-351        #domain transmembrane #status predicted #label TW6\
SUMMARY        #length 457 #molecular-weight 51822 #checksum 5700

Query Match      77.3%; Score 51; DB 2; Length 457;
Best Local Similarity 71.4%; Pred. No. 1,14e+00;
Matches          5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 279 YLTCPE 285
||:||||
Qy 531 YLSCPM 537

RESULT 4
ENTRY   #type complete
TITLE   hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) -
              sea urchin (Strongylocentrotus purpuratus)
ORGANISM #formal_name Strongylocentrotus purpuratus #common_name
              purple urchin
DATE      21-May-1990 #sequence_revision 21-May-1990 #text_change
              20-Mar-1998
ACCESSIONS A31898; A28367
REFERENCE  A31898
#authors   Woodward, H.D.; Allen, J.M.C.; Lennarz, W.J.
#journal   J. Biol. Chem. (1988) 263:18411-18418
#title     3-hydroxy-3-methylglutaryl-coenzyme A reductase of the sea
              urchin embryo. Deduced structure and regulatory properties.
#cross-references MUID:89054023
#accession  A31898
#molecule_type mRNA
#residues   1-932 #label WOO
#cross-references GB:J04200; NID:gl61522; PID:gl61523
#note       the authors rearranged portions of the coding region in

```

Figure 2, and the above sequence is taken directly from Figure 3: it matches the translation of the nucleotide sequence that the author submitted to Genbank

##note the authors translated the codon GGA for residue 805 as Glu

REFERENCE A28367
#authors Woodward, H.D.; Allen, J.M.C.; Lennarz, W.J.
#journal J. Biol. Chem. (1988) 263:2513-2517
#title 3-hydroxy-3-methylglutaryl coenzyme A reductase in the sea urchin embryo is developmentally regulated.

#cross-references MUID:88115403
#accession A28367
#molecule_type mRNA
#residues 689-735 #label WO2
COMMENT This transmembrane glycoprotein of the endoplasmic reticulum is involved in the control of cholesterol biosynthesis.

CLASSIFICATION #superfamily hydroxymethylglutaryl-CoA reductase (NADPH)
KEYWORDS cholesterol biosynthesis; endoplasmic reticulum; glycoprotein; NADP; oxidoreductase

FEATURE 279,850,886,930 #binding_site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 932 #molecular-weight 100965 #checksum 3584

Query Match 75.8%; Score 50; DB 2; Length 932;
Best Local Similarity 75.0%; Pred. No. 1.89e+00;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 808 YLSCMPS 815
|||||
QY 531 YLSCPMES 538

RESULT 5
ENTRY A49970
TITLE glycine receptor alpha-4 chain - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 17-Mar-1999

ACCESSIONS A49970
REFERENCE A49970
#authors Matzenbach, B.; Maulet, Y.; Sefton, L.; Courtier, B.; Avner, P.; Guenet, J.L.; Betz, H.

#journal J. Biol. Chem. (1994) 269:2607-2612
#title Structural analysis of mouse glycine receptor alpha subunit genes. Identification and chromosomal localization of a novel variant alpha4.

#cross-references MUID:94132024
#accession A49970
#status Preliminary; nucleic acid sequence not shown
#molecule_type DNA
#residues 1-337 #label MAT
#cross-references GB:X75850; NID:q435513; PID:g817957

GENETICS Gla4
#gene #superfamily acetylcholine receptor
CLASSIFICATION neurotransmitter receptor; transmembrane protein
KEYWORDS #length 337 #checksum 9603

Query Match 74.2%; Score 49; DB 2; Length 337;
Best Local Similarity 83.3%; Pred. No. 3.12e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 146 LSCPM 151
|||||
QY 532 LSCPME 537

RESULT 6
ENTRY S34469
TITLE gamma-aminobutyric acid receptor-like protein - great pond snail

ORGANISM #formal_name Lymnaea stagnalis #common_name great pond snail
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Mar-1999

ACCESSIONS S34469
REFERENCE S34469
#authors Hutton, M.L.; Harvey, R.J.; Earley, F.G.P.; Barnard, E.A.; Darlison, M.G.

#journal FEBS Lett. (1993) 326:112-116
#title A novel invertebrate GABA(A) receptor-like polypeptide. Sequence and pattern of gene expression.

#cross-references MUID:93314766
#accession S34469
#status Preliminary
#molecule_type mRNA
#residues 1-437 #label HUT
#cross-references EMBL:X71357; NID:g396715; PID:g396716
CLASSIFICATION #superfamily acetylcholine receptor
KEYWORDS neurotransmitter receptor
SUMMARY #length 437 #molecular-weight 50013 #checksum 1643

Query Match 74.2%; Score 49; DB 2; Length 437;
Best Local Similarity 83.3%; Pred. No. 3.12e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 164 LSCPMD 169
|||||
QY 532 LSCPME 537

RESULT 7
ENTRY B49970
TITLE glycine receptor alpha-2 chain - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 17-Mar-1999

ACCESSIONS B49970; S15802
REFERENCE A49970
#authors Matzenbach, B.; Maulet, Y.; Sefton, L.; Courtier, B.; Avner, J.; Guenet, J.L.; Betz, H.

#journal J. Biol. Chem. (1994) 269:2607-2612
#title Structural analysis of mouse glycine receptor alpha subunit genes. Identification and chromosomal localization of a novel variant alpha4.

#cross-references MUID:94132024
#accession B49970
#status Preliminary; nucleic acid sequence not shown
#molecule_type DNA
#residues 1-451 #label MAT
#cross-references GB:X75841

REFERENCE S15776
#authors Kuhse, J.; Kuryatov, A.; Maulet, Y.; Malosio, M.L.; Schmieden, V.; Betz, H.

#journal FEBS Lett. (1991) 283:73-77
#title Alternative splicing generates two isoforms of the alpha-2 subunit of the inhibitory glycine receptor.

#cross-references MUID:91243883
#accession S15802
#status not compared with conceptual translation
#molecule_type mRNA
#residues 67-89 #label KUH

GENETICS Glra2
#gene #superfamily acetylcholine receptor
CLASSIFICATION neurotransmitter receptor; transmembrane protein
KEYWORDS #length 451 #molecular-weight 51933 #checksum 263

Query Match 74.2%; Score 49; DB 2; Length 451;
Best Local Similarity 83.3%; Pred. No. 3.12e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 169 LSCPMD 174
|||||
QY 532 LSCPME 537

```

8
RESULT      8
ENTRY       S12381      #type complete
TITLE       glycine receptor alpha-2 chain - human
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
ACCESSIONS  S12381
REFERENCE   S12381
#authors    Grenningloh, G.; Schmieden, V.; Schofield, P.R.; Seeburg,
#journal     P.H.; Siddique, T.; Mohandas, T.K.; Becker, C.M.; Betz, H.
#title       EMBO J. (1990) 9:771-776
#cross-references MUID:90183975
#accession  S12381
#status     preliminary
#molecule_type mRNA
#residues   1-452 #label GRE
#cross-references GB:X52008; NID:g31848; PID:g31849
#note       the sequence from Fig. 2 is inconsistent with that from
              Fig. 1 in lacking 16-Phe
CLASSIFICATION #superfamily acetylcholine receptor
KEYWORDS       transmembrane protein
SUMMARY        #length 452 #molecular-weight 52002 #checksum 5593

Query Match      74.2%; Score 49; DB 2; Length 452;
Best Local Similarity 83.3%; Pred. No. 3.12e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 170 LSCPMD 175
QY 532 LSCPME 537
|||||:

RESULT      9
ENTRY       S14816      #type complete
TITLE       glycine receptor alpha-2 chain variant A precursor - rat
ALTERNATE_NAMES neonatal glycine receptor
ORGANISM      #formal_name Rattus norvegicus #common_name Norway rat
DATE          30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
ACCESSIONS    S14816; S15776; JN0112
REFERENCE     S14816
#authors      Akagi, H.; Hirai, K.; Hishinuma, F.
#journal      FEBS Lett. (1991) 281:160-166
#title        Cloning of a glycine receptor subtype expressed in rat brain
              and spinal cord during a specific period of neuronal
              development.
#cross-references MUID:91200276
#accession    S14816
#molecule_type DNA
#residues     1-452 #label AKA
#cross-references EMBL:X57281; NID:g56743; PID:g56744
REFERENCE     S15776
#authors      Kuhse, J.; Kuryatov, A.; Maulet, Y.; Malosio, M.L.;
              Schmieden, V.; Betz, H.
#journal      FEBS Lett. (1991) 283:73-77
#title        Alternative splicing generates two isoforms of the alpha-2
              subunit of the inhibitory glycine receptor.
#cross-references MUID:91243883
#accession    S15776
#status       not compared with conceptual translation
#molecule_type mRNA
#residues     1-452 #label KUH
#cross-references GB:X61159; NID:g288344; PID:g288345
REFERENCE     JN0112
#authors      Kuhse, J.; Schmieden, V.; Betz, H.
#journal      Neuron (1990) 5:867-873
#title        A single amino acid exchange alters the pharmacology of
              neonatal rat glycine receptor subunit.

#cross-references MUID:91097798
#accession    JN0112
#molecule_type mRNA
#residues     1-193,'E',195-452 #label KU2
COMMENT       Glycine reduces neuronal firing by activating this inhibitory
              glycine receptor, a chloride channel protein abundant in spinal
              cord and brainstem of vertebrates.
CLASSIFICATION #superfamily acetylcholine receptor
KEYWORDS       glycoprotein; ion channel; neurotransmitter receptor;
              transmembrane protein
FEATURE       #domain signal sequence #status predicted #label SIG\
              1-27
              28-452
              #product glycine receptor alpha-2 chain #status
              predicted #label MAT\
              #domain transmembrane #status predicted #label TM1\
              253-280
              287-304
              #domain transmembrane #status predicted #label TM2\
              316-342
              #domain transmembrane #status predicted #label TM3\
              424-441
              #domain transmembrane #status predicted #label TM4\
              72,103
              #binding_site carbohydrate (Asn) (covalent) #status
              predicted
SUMMARY        #length 452 #molecular-weight 51981 #checksum 4906

Query Match      74.2%; Score 49; DB 2; Length 452;
Best Local Similarity 83.3%; Pred. No. 3.12e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 170 LSCPMD 175
QY 532 LSCPME 537
|||||:

RESULT      10
ENTRY       S18836      #type complete
TITLE       glycine receptor alpha-2 chain precursor variant B - rat
ORGANISM      #formal_name Rattus norvegicus #common_name Norway rat
DATE          19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
              21-Aug-1998
ACCESSIONS    S18836
REFERENCE     S15776
#authors      Kuhse, J.; Kuryatov, A.; Maulet, Y.; Malosio, M.L.;
              Schmieden, V.; Betz, H.
#journal      FEBS Lett. (1991) 283:73-77
#title        Alternative splicing generates two isoforms of the alpha-2
              subunit of the inhibitory glycine receptor.
#cross-references MUID:91243883
#accession    S18836
#status       preliminary
#molecule_type mRNA
#residues     1-452 #label KUH
CLASSIFICATION #superfamily acetylcholine receptor
KEYWORDS       glycoprotein; membrane protein
FEATURE       #domain signal sequence #status predicted #label SIG\
              1-27
              28-452
              #product glycine receptor alpha-2 chain variant B
              #status predicted #label MAT\
              #binding_site carbohydrate (Asn) (covalent) #status
              predicted
SUMMARY        #length 452 #molecular-weight 51965 #checksum 4375

Query Match      74.2%; Score 49; DB 2; Length 452;
Best Local Similarity 83.3%; Pred. No. 3.12e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 170 LSCPMD 175
QY 532 LSCPME 537
|||||:

RESULT      11
ENTRY       A23682      #type complete
TITLE       glycine receptor alpha-3 chain precursor - rat
ORGANISM      #formal_name Rattus norvegicus #common_name Norway rat
DATE          04-Oct-1991 #sequence_revision 04-Oct-1991 #text_change

```



```

21-Aug-1998
ACCESSIONS A23682
REFERENCE A23682
#authors Kuhse, J.; Schmieden, V.; Betz, H.
#journal J. Biol. Chem. (1990) 265:22317-22320
#title Identification and functional expression of a novel ligand
binding subunit of the inhibitory glycine receptor.
#cross-references MUID:91093073
#accession A23682
##status preliminary
##molecule_type mRNA
##residues 1-464 ##label KUH
##cross-references GB:M5250; GB:M38385; NID:9204882; PID:9204883
CLASSIFICATION #superfamily acetylcholine receptor
KEYWORDS transmembrane protein
SUMMARY #length 464 #molecular-weight 53672 #checksum 4347

Query Match 74.2%; Score 49; DB 2; Length 464;
Best Local Similarity 83.3%; Pred. No. 3.12e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 169 LSCPM 174
Qy 532 LSCPM 537

RESULT 12
ENTRY S39554 #type complete
TITLE histidine decarboxylase (EC 4.1.1.22) - tomato
ORGANISM #formal_name Lycopersicon esculentum #common_name tomato
DATE 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
05-Dec-1998
ACCESSIONS S39554
REFERENCE S39554
#authors Picton, S.; Gray, J.E.; Payton, S.; Barton, S.L.; Lowe, A.;
Grierson, D.
#journal Plant Mol. Biol. (1993) 23:627-631
#title A histidine decarboxylase-like mRNA is involved in tomato
fruit ripening.
#cross-references MUID:94033342
#accession S39554
##molecule_type mRNA
##residues 1-413 ##label PIC
##cross-references EMBL:X7190; NID:9416533; PID:9416534
#note the authors did not translate the codon for residue 33
CLASSIFICATION #superfamily Kiebsiella histidine decarboxylase
KEYWORDS carbon-carbon lyase; carboxy-lyase; phosphoprotein; pyridoxal
phosphate
FEATURE 242
#binding_site pyridoxal phosphate (Lys) (covalent)
#status predicted
SUMMARY #length 413 #molecular-weight 47628 #checksum 7578

Query Match 72.7%; Score 48; DB 1; Length 413;
Best Local Similarity 66.7%; Pred. No. 5.12e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 243 FLGCPM 248
Qy 531 YLSCPM 536

RESULT 13
ENTRY A70988 #type complete
TITLE hypothetical protein Rv1760 - Mycobacterium tuberculosis
(strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS A70988
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry

```

```

III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagers, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skilton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession A70988
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-502 ##label COL
##cross-references GB:295890; NID:93242245; PID:e319004;
PID:g2131028
##experimental_source strain H37Rv
GENETICS
#gene Rv1760
SUMMARY #length 502 #molecular-weight 54092 #checksum 4814

Query Match 72.7%; Score 48; DB 2; Length 502;
Best Local Similarity 50.0%; Pred. No. 5.12e+00;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 427 FLGCRMS 434
Qy 531 YLSCPM 538

RESULT 14
ENTRY S46347 #type complete
TITLE pol polyprotein - simian immunodeficiency virus SIVagm
( isolate SAB-1)
ORGANISM #formal_name simian immunodeficiency virus SIVagm
#variety isolate SAB-1
DATE 25-Dec-1994 #sequence_revision 14-Feb-1997 #text_change
17-Mar-1999
ACCESSIONS S46347
REFERENCE S46335
#authors Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.;
Barre-Sinoussi, F.; Hirsch, V.M.; Allan, J.S.; Shaw, G.M.;
Sharp, P.M.; Hahn, B.H.
#journal EMBO J. (1994) 13:2935-2947
#title Mosaic genome structure of simian immunodeficiency virus from
West African green monkeys.
#cross-references MUID:94298785
#accession S46347
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-1039 ##label JIN
##cross-references EMBL:U04005; NID:9466229; PID:g466231
#experimental_source isolate SAB-1; sabaeus monkey
#note the nucleotide sequence was submitted to the EMBL Data
Library, December 1993
#note this reading frame extends between two stop codons and
does not begin with a start codon
GENETICS
#gene pol
CLASSIFICATION #superfamily pol polyprotein
SUMMARY #length 1039 #molecular-weight 117966 #checksum 3256

Query Match 72.7%; Score 48; DB 2; Length 1039;
Best Local Similarity 50.0%; Pred. No. 5.12e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 302 YFSCP 309
Qy 531 YLSCPM 538

```

```

RESULT 15
ENTRY S78727 #type complete
TITLE protein YLL018c-a - yeast (Saccharomyces cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 15-Jan-1999 #sequence_revision 15-Jan-1999 #text_change
15-Jan-1999
ACCESSIONS S78727
REFERENCE S64761
#authors Goffeau, A.; Purnelle, B.
#submission submitted to the Protein Sequence Database, May 1996
#accession S78727
##status preliminary
##molecule_type DNA
##residues 1-98 #label GOF
##cross-references EMBL:Z73125; MIPS:YLL018c-a
##experimental_source strain S288C
GENETICS
#map_position 12L
SUMMARY #length 98 #molecular-weight 11105 #checksum 9393
Query Match 71.2%; Score 47; DB 2; Length 98;
Best Local Similarity 57.1%; Pred. No. 8.34e+00;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 59 YLRCRMD 65
Qy 531 YLSCPME 537
Search completed: Thu Jul 8 19:02:38 1999
Job time : 11 secs.

```

WQREH (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:01:32 1999; MasPar time 4.24 Seconds
Tabular output not generated. 53.397 Million cell updates/sec

Title: >US-09-041-236-2
Description: (531-538) from US09041236.pep (27 of 45)
Perfect Score: 66
Sequence: 1 YLSCPMS 8

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 22.515; Variance 24.352; scale 0.925

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	53	80.3	GID_BACSU	GID PROTEIN.	1.50e-01
2	51	77.3	CD51_YEAST	PHOSPHATIDATE CYTIDYLY	4.55e-01
3	51	77.3	DCA1_METFR	CARBON MONOXIDE DEHYDR	4.55e-01
4	50	75.8	HMDH_STRPU	3-HYDROXY-3-METHYLGLUT	7.84e-01
5	49	74.2	GRA4_MOUSE	GLYCINE RECEPTOR ALPHA	1.34e+00
6	49	74.2	GRA2_RAT	GLYCINE RECEPTOR ALPHA	1.34e+00
7	49	74.2	GRA2_HUMAN	GLYCINE RECEPTOR ALPHA	1.34e+00
8	49	74.2	GRA3_RAT	GLYCINE RECEPTOR ALPHA	1.34e+00
9	49	74.2	GRA3_HUMAN	GLYCINE RECEPTOR ALPHA	1.34e+00
10	48	72.7	YEB7_SCHPO	HYPOTHETICAL 11.3 KD P	2.28e+00
11	48	72.7	DCHS_LYCCE	HISTIDINE DECARBOXYLAS	2.28e+00
12	47	71.2	GRA1_MOUSE	GLYCINE RECEPTOR ALPHA	3.85e+00
13	47	71.2	GRA1_HUMAN	GLYCINE RECEPTOR ALPHA	3.85e+00
14	47	71.2	GRA1_RAT	GLYCINE RECEPTOR ALPHA	3.85e+00
15	47	71.2	YJB3_YEAST	HYPOTHETICAL 59.5 KD P	3.85e+00
16	47	71.2	POL_SIVCZ	POL POLYPROTEIN [CONTA	3.85e+00
17	46	69.7	F26L_MOUSE	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
18	46	69.7	F26B_HUMAN	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
19	46	69.7	YAKA_SCHPO	HYPOTHETICAL 25.4 KD P	6.44e+00
20	46	69.7	YGS6_YEAST	HYPOTHETICAL 29.4 KD P	6.44e+00
21	46	69.7	IL1B_HORSE	INTERLEUKIN-1 BETA PRE	6.44e+00
22	46	69.7	SVH2_SYNV3	PROBABLE HISTIDYL-TRNA	6.44e+00
23	46	69.7	F262_RAT	6-PHOSPHOFRUCTO-2-KINA	6.44e+00

24	46	69.7	1	F26B_BOVIN	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
25	46	69.7	1	F263_RAT	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
26	46	69.7	1	F26L_CHICK	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
27	46	69.7	1	F26L_RANCA	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
28	46	69.7	1	F26L_RAT	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
29	46	69.7	1	F26L_HUMAN	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
30	46	69.7	1	F26L_BOVIN	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
31	46	69.7	1	C7B3_ARATH	CYTOCHROME P450 71B3 (6.44e+00
32	46	69.7	1	F26P_HUMAN	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
33	46	69.7	1	F26H_BOVIN	6-PHOSPHOFRUCTO-2-KINA	6.44e+00
34	46	69.7	1	API_SCHPO	AP-1-LIKE TRANSCRIPTO	6.44e+00
35	46	69.7	1	RPOC_CHLVU	DNA-DIRECTED RNA POLYM	6.44e+00
36	45	68.2	1	IL1B_PIG	INTERLEUKIN-1 BETA PRE	1.07e+01
37	45	68.2	1	IL1B_FELCA	INTERLEUKIN-1 BETA PRE	1.07e+01
38	45	68.2	1	IL1B_RABIT	INTERLEUKIN-1 BETA PRE	1.07e+01
39	45	68.2	1	IL1B_RAT	INTERLEUKIN-1 BETA PRE	1.07e+01
40	45	68.2	1	IL1B_MOUSE	INTERLEUKIN-1 BETA PRE	1.07e+01
41	45	68.2	1	DUS7_RAT	DUAL SPECIFICITY PROTE	1.07e+01
42	45	68.2	1	DUS7_HUMAN	DUAL SPECIFICITY PROTE	1.07e+01
43	45	68.2	1	DUS6_RAT	DUAL SPECIFICITY PROTE	1.07e+01
44	45	68.2	1	PBG1_HUMAN	PREGNANCY-SPECIFIC BET	1.07e+01
45	45	68.2	1	EPAL_HUMAN	EPHRIN TYPE-A RECEPTOR	1.07e+01

ALIGNMENTS

RESULT 1
ID GID_BACSU STANDARD; PRT; 435 AA.
AC P39815;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GID PROTEIN.
GN GID.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA FOULGER D., ERRINGTON J.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 1-363 FROM N.A.
RC STRAIN=168 / 8G5;
RA DE JONG S.;
RL SUBMITTED (XXX-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ000975; E332181; -
CC EMBL; L27797; G520754; -
DR SUBTILIST; BG11008; GID.
DR PROSITE; PS01280; GIDA_1; FALSE_NEG.
DR PROSITE; PS01281; GIDA_2; 1.
SQ SEQUENCE 435 AA; 48063 MW; 2C2D754A CRC32;
Query Match 80.3%; Score 53; DB 1; Length 435;
Best Local Similarity 83.3%; Pred. No. 1.50e-01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 189 YLNCMP 194

Qy 531 YLSCPM 536

RESULT 2
ID CDS1_YEAST STANDARD; PRT; 457 AA.
AC P38221;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PHOSPHATIDATE CYTIDYLTRANSFERASE (EC 2.7.7.41) (CDP-DIGLYCERIDE
DE SYNTHETASE) (CDP-DIGLYCERIDE PYROPHOSPHORYLASE) (CDP-DIACYLGLYCEROL
DE SYNTHASE) (CDS) (CTP:PHOSPHATIDATE CYTIDYLTRANSFERASE) (CDP-DAG
DE SYNTHASE).
GN CDS1 OR CDG1 OR YBR029C OR YBR0313.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETEALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE; 94378725.
RA SMITS P.H.M., DE HAAN M., MAAT C., GRIVELL L.A.;
RT "The complete sequence of a 33 kb fragment on the right arm of
RT chromosome II from *Saccharomyces cerevisiae* reveals 16 open reading
RT frames, including ten new open reading frames, five previously
RT identified genes and a homologue of the *SCD1* gene.",
RL YEAST 10:S75-S80(1994).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE; 96132809.
RA SHEN H., HEACOCK P.N., CLANCEY C.J., DOWHAN W.;
RT "The *CDS1* gene encoding CDP-diacylglycerol synthase in *Saccharomyces*
RT *cerevisiae* is essential for cell growth.",
RL J. BIOL. CHEM. 271:789-795(1996).
RN [3]
RP MUTAGENESIS OF CYS-102.
RX MEDLINE; 97067155.
RA SHEN H., DOWHAN W.;
RT "Reduction of CDP-diacylglycerol synthase activity results in the
RT excretion of inositol by *Saccharomyces cerevisiae*.",
RL J. BIOL. CHEM. 271:29043-29048(1996).
CC -!- FUNCTION: SUPPLIES CDP-DIACYLGLYCEROL, WHICH MAY PLAY AN IMPORTANT
CC ROLE AS BOTH A PRECURSOR TO PHOSPHOINOSITIDE BIOSYNTHESIS IN THE
CC PLASMA MEMBRANE AND AS A NEGATIVE EFFECTOR OF PHOSPHATIDYLINOSITOL
CC 4-KINASE ACTIVITY, THEREBY EXERTING AN EFFECT ON CELL
CC PROLIFERATION VIA A LIPID-DEPENDENT SIGNAL TRANSDUCTION CASCADE.
CC -!- CATALYTIC ACTIVITY: CTP + PHOSPHATIDATE = PYROPHOSPHATE +
CC CDP-DIACYLGLYCEROL.
CC -!- PATHWAY: PHOSPHOLIPID BIOSYNTHESIS.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC MITOCHONDRION AND ENDOPLASMIC RETICULUM.
CC -!- SIMILARITY: BELONGS TO THE CDS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z35898; G536240; -.
CC DR EMBL; X76078; G498757; -.
CC PIR: S45885; S45885.
CC SGD: L0002673; CDS1.
CC DR PROSITE: PS01315; CDS; 1.
CC PFAM: PF01148; Cytidylyltrans; 1.
CC TRANSFERASE; NUCLEOTIDYLTRANSFERASE; PHOSPHOLIPID BIOSYNTHESIS;
CC TRANSFERASE; MITOCHONDRION; ENDOPLASMIC RETICULUM.
FT TRANSMEM 71 91
FT TRANSMEM 154 174
FT TRANSMEM 188 208
FT TRANSMEM 214 234
FT TRANSMEM 255 275
FT TRANSMEM 330 350
FT TRANSMEM 350 350

FT MUTAGEN 102 102 C->Y: REDUCED ENZYME LEVEL.
SQ SEQUENCE 457 AA; 51823 MW; B5162BB4 CRC32;
Query Match 77.3%; Score 51; DB 1; Length 457;
Best Local Similarity 71.4%; Pred. No. 4.55e-01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 279 YLCPVE 285
YLCPVE 531 YLCPWE 537
RESULT 3
ID DCAL_METFR STANDARD; PRT; 805 AA.
AC Q49161;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CARBON MONOXIDE DEHYDROGENASE I ALPHA SUBUNIT (EC 1.2.99.2).
GN CDH1A.
OS METHANOSARCINA FRISIA.
OC ARCHAEA; EURYARCHAEOTA; METHANOMICROBIALES; METHANOSARCINACEAE;
OC METHANOSARCINA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC STRAIN=GOEI;
RX MEDLINE; 96278885.
RA EGGEN R.I.L., VAN KRANENBURG R., VRIESEMA A.J.M., GEERLING A.C.M.,
RA VERHAGEN M.F.J.M., HAGEN W.R., DE VOS W.M.;
RT "Carbon monoxide dehydrogenase from *Methanosarcina frisia* Gol.
RT Characterization of the enzyme and the regulated expression of two
RT operon-like *cdh* gene clusters.",
RL J. BIOL. CHEM. 271:14256-14263(1996).
CC -!- FUNCTION: CATALYZES THE CLEAVAGE OF ACETYL-COA. OPTIMUM ACTIVITY
CC IS MEASURED BETWEEN PH 8 AND 9.
CC -!- CATALYTIC ACTIVITY: CO + H(2)O + ACCEPTOR -> CO(2) + REDUCED
CC ACCEPTOR.
CC -!- COFACTOR: TWO NICKEL AND FOUR IRON-SULFUR CLUSTERS PER ENZYME
CC (PROBABLY TWO 4FE-4S AND TWO 6FE-6S).
CC -!- SUBUNIT: HOMOTETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC -!- SIMILARITY: TO M.SDHENGENII CARBON MONOXIDE DEHYDROGENASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L26487; G609518; -.
CC DR PROSITE: PS00198; 4FBAS_FERREDOXIN; 1.
CC PFAM: PF00037; fer4; 1.
CC HSP: P00195; ICLF.
CC KW OXIDOREDUCTASE; NICKEL; IRON-SULFUR; 4FE-4S.
FT INIT_MET 0 0
FT METAL 75 75 IRON-SULFUR (6FE-6S) (BY SIMILARITY).
FT METAL 78 78 IRON-SULFUR (6FE-6S) (BY SIMILARITY).
FT METAL 93 93 IRON-SULFUR (6FE-6S) (BY SIMILARITY).
FT METAL 416 416 IRON-SULFUR (BY SIMILARITY).
FT METAL 419 419 IRON-SULFUR (BY SIMILARITY).
FT METAL 422 422 IRON-SULFUR (BY SIMILARITY).
FT METAL 426 426 IRON-SULFUR (BY SIMILARITY).
FT METAL 454 454 IRON-SULFUR (BY SIMILARITY).
FT METAL 457 457 IRON-SULFUR (BY SIMILARITY).
FT METAL 460 460 IRON-SULFUR (BY SIMILARITY).
FT METAL 464 464 IRON-SULFUR (BY SIMILARITY).
SQ SEQUENCE 805 AA; 88451 MW; 41FEB971 CRC32;
Query Match 77.3%; Score 51; DB 1; Length 805;
Best Local Similarity 71.4%; Pred. No. 4.55e-01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 423 YLACPIE 429
||:|:|:|
QY 531 YLSCPME 537

RESULT 4
ID HMDH_STRPU STANDARD; PRT; 932 AA.
AC P16393;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE 3-HYDROXY-3-METHYLGUTARYL-COENZYME A REDUCTASE (EC 1.1.1.34) (HMG-COA
REDUCTASE).
OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC STRONGYLOCENTROTUS.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 8905403.
RA WOODWARD H.D., ALLEN J.M.C., LENNARZ W.J.;
RT "3-hydroxy-3-methylglutaryl-coenzyme A reductase of the sea urchin
embryo. Deduced structure and regulatory properties.";
RL J. BIOL. CHEM. 263:18411-18418(1988).
RN [2]
RN SEQUENCE OF 689-735 FROM N.A.
RX MEDLINE; 88115403.
RA WOODWARD H.D., ALLEN J.M.C., LENNARZ W.J.;
RT "3-hydroxy-3-methylglutaryl coenzyme A reductase in the sea urchin
embryo is developmentally regulated.";
RL J. BIOL. CHEM. 263:2513-2517(1988).
CC -!- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE
CC CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING
CC ENZYME OF STEROL BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: (R)-MEVALONATE + COA + 2 NADP(+) = (S)-3-
CC HYDROXY-3-METHYLGUTARYL-COA + 2 NADPH.
CC -!- PATHWAY: CHOLESTEROL BIOSYNTHESIS
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -!- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J04200; G161523; -;
CC EMBL; J03523; -; NOT_ANNOTATED_CDS.
CC PIR; A28367; A28367.
CC PIR; A31898; A31898.
CC PROSITE; PS00066; HMG_COA_REDUCTASE_1; 1.
CC PROSITE; PS00318; HMG_COA_REDUCTASE_2; 1.
CC PROSITE; PS01192; HMG_COA_REDUCTASE_3; 1.
CC PROSITE; PS50065; HMG_COA_REDUCTASE_4; 1.
CC PFAM; PF00368; HMG-COA_red1; 1.
CC PFAM; PF00369; HMG-COA_red2; 1.
CC OXIDOREDUCTASE; GLYCOPROTEIN; ENDOPLASMIC RETICULUM; TRANSMEMBRANE;
KW CHOLESTEROL BIOSYNTHESIS; NADP.
FT DOMAIN 1 338 MEMBRANE-BOUND.
FT DOMAIN 339 467 LINKER.
FT DOMAIN 468 932 CATALYTIC.
FT ACT_SITE 575 575 BY SIMILARITY.
FT ACT_SITE 783 783 BY SIMILARITY.
FT ACT_SITE 882 882 GENERAL BASE (BY SIMILARITY).
FT CARBOHYD 279 279 POTENTIAL.
FT CARBOHYD 850 850 POTENTIAL.
FT CARBOHYD 886 886 POTENTIAL.
SQ SEQUENCE 932 AA; 100965 MW; 859F4C02 CRC32;

Query Match 75.88; Score 50; DB 1; Length 932;

Best Local Similarity 75.08; Pred. No. 7.84e-01;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 808 YLSCTMPs 815
|||||
QY 531 YLSCPMEs 538

RESULT 5
ID GRAA_MOUSE STANDARD; PRT; 337 AA.
AC Q61603;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLYCINE RECEPTOR ALPHA-4 CHAIN PRECURSOR (FRAGMENT).
GN GLRA4.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN-BALB/C;
RX MEDLINE; 94132024.
RA MATZENBACH B., MAULET Y., SEFTON L., COURTIER B., AVNER P.,
RA CUENET J.-L., BETZ H.;
RT Structural analysis of mouse glycine receptor alpha subunit genes.
RT Identification and chromosomal localization of a novel variant.";
RL J. BIOL. CHEM. 269:2607-2612(1994).
CC -!- FUNCTION: THE GLYCINE RECEPTOR IS A NEUROTRANSMITTER-GATED ION
CC CHANNEL. BINDING OF GLYCINE TO ITS RECEPTOR INCREASES THE CHLORIDE
CC CONDUCTANCE AND THUS PRODUCES HYPERPOLARIZATION (INHIBITION OF
CC NEURONAL FIRING).
CC -!- SUBUNIT: PENTAMER COMPOSED OF ALPHA (48 KD) AND BETA (58 KD)
CC SUBUNITS. A THIRD TYPE OF SUBUNIT (93 KD) IS A PERIPHERAL MEMBRANE
CC PROTEIN ASSOCIATED WITH THE CYTOPLASMIC DOMAINS OF THE RECEPTOR
CC CORE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- THE ALPHA SUBUNIT BINDS STRYCHNINE.
CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X75850; G817957; -;
CC EMBL; X75851; G817957; JOINED.
CC EMBL; X75852; G817957; JOINED.
CC EMBL; X75853; G817957; JOINED.
CC MGD; MGI:95750; GLRA4.
CC PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
CC PFAM; PF00065; neur_chan; 1.
CC RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
KW TRANSMEMBRANE; MULTIGENE FAMILY.
FT NON_TER 1 1
FT SIGNAL <1 10 POTENTIAL.
FT CHAIN 11 >337 GLYCINE RECEPTOR ALPHA-4 SUBUNIT.
FT DOMAIN 11 234 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
FT DOMAIN 315 >337 CYTOPLASMIC (PROBABLE).
FT DISULFID 148 162 BY SIMILARITY.
FT DISULFID 209 220 BY SIMILARITY.
FT CARBOHYD 48 48 POTENTIAL.
FT NON_TER 337 337
SQ SEQUENCE 337 AA; 6CABAD4D CRC32;

Query Match 74.28; Score 49; DB 1; Length 337;
Best Local Similarity 83.38; Pred. No. 1.34e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Db 146 LSCPM 151
Cc |||||:
Qy 532 LSCPME 537

RESULT 6
ID GRA2-RAT STANDARD; PRT; 452 AA.
AC P22771;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE (GLYCINE RECEPTOR ALPHA-2* CHAIN PRECURSOR (NEONATAL ISOFORM)
GN GLRA2.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ROENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RX MEDLINE: 91097798.
RA KUHSE J., SCHMIEDEN V., BETZ H.;
RT "A single amino acid exchange alters the pharmacology of neonatal rat
glycine receptor subunit.";
RL NEURON 5:867-873(1990).
RN [2]
RX SEQUENCE FROM N.A.
RC STRAIN WISTAR; TISSUE-SPINAL CORD;
RX MEDLINE: 91200276.
RA AKAGI H., HIRAI K., HISHINUMA F.;
RT "Cloning of a glycine receptor subtype expressed in rat brain and
spinal cord during a specific period of neuronal development.";
RL FEBS LETT. 281:160-166(1991).
RN [3]
RX SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA KUHSE J., KURYATOV A., MAULET Y., MALOSIO M.L., SCHMIEDEN V.,
RA BETZ H.;
RT "Alternative splicing generates two isoforms of the alpha 2 subunit
of the inhibitory glycine receptor.";
RL FEBS LETT. 283:73-77(1991).
CC -!- FUNCTION: THE GLYCINE RECEPTOR IS A NEUROTRANSMITTER-GATED ION
CHANNEL. BINDING OF GLYCINE TO ITS RECEPTOR INCREASES THE CHLORIDE
CONDUCTANCE AND THUS PRODUCES HYPERPOLARIZATION (INHIBITION OF
NEURONAL FIRING).
CC -!- SUBUNIT: PENTAMER COMPOSED OF ALPHA (48 KD) AND BETA (58 KD)
SUBUNITS. A THIRD TYPE OF SUBUNIT (93 KD) IS A PERIPHERAL MEMBRANE
PROTEIN ASSOCIATED WITH THE CYTOPLASMIC DOMAINS OF THE RECEPTOR
CORE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: THE GLYCINE RECEPTOR IS ABUNDANT IN THE SPINAL
CORD AND BRAINSTEM OF VERTEBRATES.
CC -!- DEVELOPMENTAL STAGE: THE ALPHA-2* SUBUNIT ISOFORM IS PRESENT ONLY
IN NEONATAL RATS. ISOFORMS ALPHA-2A AND ALPHA-2B ARE PRESENT IN
THE PRE- AND NEONATAL BRAIN. AT LATER POSTNATAL STAGES, ALPHA-2A
LEVELS GREATLY DECREASE WHILE ALPHA-2B IS BARELY DETECTABLE.
CC -!- THE ALPHA SUBUNIT BINDS STRYCHNINE.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS OF THE ALPHA-2 SUBUNIT (ALPHA-2*
AND ALPHA-2B) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME
GENE.
CC -!- SIMILARITY: IDENTICAL TO THE HUMAN ALPHA-2 SUBUNIT, EXCEPT FOR 5
SUBSTITUTIONS AT POSITIONS 18, 24, 37, 194 AND 404. SUBSTITUTION
AT POSITION 194 (G -> E) ACCOUNTS FOR THE LOWER STRYCHNINE
SENSITIVITY OBSERVED IN NEONATAL RATS.
CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

Cc or send an email to license@isb-sib.ch).
Cc -----
Cc EMBL: X57281; G56744; -.
Cc EMBL: X61159; G288345; -.
Cc PIR: JN0112; JN0112.
Cc PIR: S14816; S14816.
Cc DR PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
Cc DR PFAM; PF00065; neur_chan; 2.
Cc KW RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
Cc KW TRANSMEMBRANE; MULTIGENE FAMILY; ALTERNATIVE SPLICING.
Cc FT SIGNAL 1 27
Cc FT CHAIN 28 452
Cc FT DOMAIN 28 253
Cc FT TRANSMEM 254 280
Cc FT TRANSMEM 287 304
Cc FT TRANSMEM 319 342
Cc FT DOMAIN 343 423
Cc FT TRANSMEM 424 441
Cc FT DISULFID 172 186
Cc FT DISULFID 232 243
Cc FT CARBOHYD 72 72
Cc FT CARBOHYD 103 103
Cc FT BINDING 194 194
Cc FT BINDING 85 86
Cc FT VARSPLIC 194 194
Cc FT VARIANT 194 194
Cc SQ SEQUENCE 452 AA; 52053 MW; 4F9B9D77 CRC32;
Cc
Cc Query Match 74.2%; Score 49; DB 1; Length 452;
Cc Best Local Similarity 83.3%; Pred. No. 1.34e+00;
Cc Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Cc
Cc Db 170 LSCPM 175
Cc |||||:
Cc Qy 532 LSCPME 537

RESULT 7
ID GRA2-HUMAN STANDARD; PRT; 452 AA.
AC P23416; Q99862;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLYCINE RECEPTOR ALPHA-2 CHAIN PRECURSOR.
GN GLRA2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 90183975.
RA GREENINGLOH G., SCHMIEDEN V., SCHOFIELD P.R., SEEBURG P.H.,
RA SIDDIQUE T., MOHANDAS T.K., BECKER C.M., BETZ H.;
RT "Alpha subunit variants of the human glycine receptor: primary
structures, functional expression and chromosomal localization of the
corresponding genes.";
RL EMBO J. 9:771-776(1990).
RN [2]
RX SEQUENCE OF 1-360 FROM N.A.
RA MONANI U.R., BURGHES A.H.M.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: THE GLYCINE RECEPTOR IS A NEUROTRANSMITTER-GATED ION
CHANNEL. BINDING OF GLYCINE TO ITS RECEPTOR INCREASES THE CHLORIDE
CONDUCTANCE AND THUS PRODUCES HYPERPOLARIZATION (INHIBITION OF
NEURONAL FIRING).
CC -!- SUBUNIT: PENTAMER COMPOSED OF ALPHA (48 KD) AND BETA (58 KD)
SUBUNITS. A THIRD TYPE OF SUBUNIT (93 KD) IS A PERIPHERAL MEMBRANE
PROTEIN ASSOCIATED WITH THE CYTOPLASMIC DOMAINS OF THE RECEPTOR
CORE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- THE ALPHA SUBUNIT BINDS STRYCHNINE.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS OF THE ALPHA-2 SUBUNIT (ALPHA-2*
AND ALPHA-2B) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME

```

CC GENE.
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X52008; G31849; -
 CC EMBL; U77731; G1710184; -
 CC EMBL; U77724; G1710184; JOINED.
 CC EMBL; U77725; G1710184; JOINED.
 CC EMBL; U77726; G1710184; JOINED.
 CC EMBL; U77727; G1710184; JOINED.
 CC EMBL; U77728; G1710184; JOINED.
 CC EMBL; U77729; G1710184; JOINED.
 CC EMBL; U77730; G1710184; JOINED.
 CC PIR; S12381; S12381.
 CC MIM; 305990; -
 CC PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
 CC PFAM; PF00065; neur_chan; 2.
 CC RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
 CC TRANSMEMBRANE; MULTIGENE FAMILY; ALTERNATIVE SPLICING.
 CC SIGNAL 1 27
 CC CHAIN 28 452
 CC DOMAIN 28 253
 CC TRANSMEM 254 280
 CC TRANSMEM 287 304
 CC TRANSMEM 319 342
 CC DOMAIN 343 423
 CC TRANSMEM 424 441
 CC DISULFID 172 186
 CC CARBOHYD 72 72
 CC CARBOHYD 103 103
 CC VARSPLIC 85 86
 CC VT -> IA (IN ALPHA-2B).
 CC SQ SEQUENCE 452 AA; 52002 MW; 6FF4B7F3 CRC32;

Query Match 74.2%; Score 49; DB 1; Length 452;
 Best Local Similarity 83.3%; Pred. No. 1.34e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 170 LSCPM 175
 Qy 532 LSCPM 537

RESULT 8
 ID GRA3_RAT
 AC P24524;
 DT 01-MAR-1992 (REL. 21, CREATED)
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE GLYCINE RECEPTOR ALPHA-3 CHAIN PRECURSOR.
 GN GLRA3.
 OS RATTUS NORVEGICUS (RAT).
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RC TISSUE=BRAIN;
 RX MEDLINE; 91093073.
 RA KUNSE J., SCHMIEDEN V., BETZ H.;
 RT "Identification and functional expression of a novel ligand binding
 RT subunit of the inhibitory glycine receptor."
 RL J. BIOL. CHEM. 265:22317-22320(1990).
 CC -!- FUNCTION: THE GLYCINE RECEPTOR IS A NEUROTRANSMITTER-GATED ION
 CC CHANNEL. BINDING OF GLYCINE TO ITS RECEPTOR INCREASES THE CHLORIDE
 CC CONDUCTANCE AND THUS PRODUCES HYPERPOLARIZATION (INHIBITION OF
 CC NEURONAL FIRING).

CC -!- SUBUNIT: PENTAMER COMPOSED OF ALPHA (48 KD) AND BETA (58 KD)
 CC SUBUNITS. A THIRD TYPE OF SUBUNIT (93 KD) IS A PERIPHERAL MEMBRANE
 CC PROTEIN ASSOCIATED WITH THE CYTOPLASMIC DOMAINS OF THE RECEPTOR
 CC CORE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- THE ALPHA SUBUNIT BINDS STRYCHNINE.
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M55250; G204883; -
 CC PIR; A23682; A23682.
 CC PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
 CC PFAM; PF00065; neur_chan; 2.
 CC RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
 CC TRANSMEMBRANE; MULTIGENE FAMILY.
 CC SIGNAL 1 33
 CC CHAIN 34 464
 CC DOMAIN 34 251
 CC TRANSMEM 253 279
 CC TRANSMEM 286 303
 CC TRANSMEM 315 341
 CC DOMAIN 342 433
 CC TRANSMEM 434 451
 CC DISULFID 171 185
 CC DISULFID 231 242
 CC CARBOHYD 71 71
 CC SQ SEQUENCE 464 AA; 53672 MW; 3DB49B79 CRC32;
 Query Match 74.2%; Score 49; DB 1; Length 464;
 Best Local Similarity 83.3%; Pred. No. 1.34e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 169 LSCPM 174
 Qy 532 LSCPM 537

RESULT 9
 ID GRA3_HUMAN
 AC O75311;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE GLYCINE RECEPTOR ALPHA-3 CHAIN PRECURSOR.
 GN GLRA3.
 OS HOMO SAPIENS (HUMAN).
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RC SEQUENCE FROM N.A.
 RX MEDLINE; 98344067.
 RA NIKOLIC Z., LAUBE B., WEBER R.G., LICHTER P., KIOSCHIS P., POUSTKA A.,
 RA MULHARDT C., BECKER C.M.;
 RT "The human glycine receptor subunit alpha3. GLRA3 gene structure,
 RT chromosomal localization, and functional characterization of
 RT alternative transcripts."
 RL J. BIOL. CHEM. 273:19708-19714(1998).
 CC -!- FUNCTION: THE GLYCINE RECEPTOR IS A NEUROTRANSMITTER-GATED ION
 CC CHANNEL. BINDING OF GLYCINE TO ITS RECEPTOR INCREASES THE CHLORIDE
 CC CONDUCTANCE AND THUS PRODUCES HYPERPOLARIZATION (INHIBITION OF
 CC NEURONAL FIRING).
 CC -!- SUBUNIT: PENTAMER COMPOSED OF ALPHA (48 KD) AND BETA (58 KD)
 CC SUBUNITS. A THIRD TYPE OF SUBUNIT (93 KD) IS A PERIPHERAL MEMBRANE
 CC PROTEIN ASSOCIATED WITH THE CYTOPLASMIC DOMAINS OF THE RECEPTOR
 CC CORE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -1- THE ALPHA SUBUNIT BINDS STRYCHNINE.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS; ALPHA3L (SHOWN HERE) AND ALPHA3K;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF017724; G3342792; -
DR EMBL; AF017715; G3342792; JOINED.
DR EMBL; AF017716; G3342792; JOINED.
DR EMBL; AF017717; G3342792; JOINED.
DR EMBL; AF017718; G3342792; JOINED.
DR EMBL; AF017719; G3342792; JOINED.
DR EMBL; AF017720; G3342792; JOINED.
DR EMBL; AF017721; G3342792; JOINED.
DR EMBL; AF017722; G3342792; JOINED.
DR EMBL; AF017723; G3342792; JOINED.
DR MIM; 600421; -
DR PROSITE; PS00236; NEUOTR_ION_CHANNEL; 1.
KW RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
KW TRANSMEMBRANE; MULTIGENE FAMILY; ALTERNATIVE SPLICING.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 465 GLYCINE RECEPTOR ALPHA-3 SUBUNIT.
FT DOMAIN 34 251 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 253 279 PROBABLE.
FT TRANSMEM 286 303 PROBABLE.
FT TRANSMEM 315 341 PROBABLE.
FT DOMAIN 342 433 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 434 451 PROBABLE.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 231 242 BY SIMILARITY.
FT CARBOHYD 71 71 POTENTIAL.
SQ SEQUENCE 465 AA; 53937 MW; C2685AA0 CRC32;

Query Match 74.2%; Score 49; DB 1; Length 465;
Best Local Similarity 83.3%; Pred. No. 1.34e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 169 LSCPM 174
QY 532 LSCPM 537
|||||

RESULT 10
ID YEB7_SCHPO STANDARD; PRT; 101 AA.
AC Q14036;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 11.3 KD PROTEIN C266.07 IN CHROMOSOME I.
GN SPAC26.07.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIZOSOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z98887; E339194; -
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 89 94 POLY-SER.
SQ SEQUENCE 101 AA; 11296 MW; 22B9D6DC CRC32;

Query Match 72.7%; Score 48; DB 1; Length 101;
Best Local Similarity 50.0%; Pred. No. 2.28e+00;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 48 YLOCMDT 55
QY 531 YLSCPM 538
|||||

RESULT 11
ID DCHS_LYCES STANDARD; PRT; 413 AA.
AC P54772;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HISTIDINE DECARBOXYLASE (EC 4.1.1.22) (HDC) (TOM92).
GN HDC.
OS LYCOPERSICON ESCULENTUM (TOMATO).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANACEAE; SOLANUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, AILSA CRAIG;
RX MEDLINE; 94033342.
RA PICTON S., GRAY J.E., PAYTON S., BARTON S.L., LOWE A., GRIERSON D.;
RT "A histidine decarboxylase-like mRNA is involved in tomato fruit
RT ripening."
RL PLANT MOL. BIOL. 23:627-631(1993).
CC -1- CATALYTIC ACTIVITY: L-HISTIDINE -> HISTAMINE + CO(2).
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- TISSUE SPECIFICITY: RIPE FRUITS; NOT DETECTED IN LEAVES AND UNRIPE
CC FRUIT.
CC -1- DEVELOPMENTAL STAGE: ACCUMULATES DURING EARLY FRUIT RIPENING AND
CC THEN DECLINES.
CC -1- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
CC TYRDC).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X71900; G416534; -
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
DR PFAM; PF00282; pyridoxal_dec; 1.
KW LYASE; DECARBOXYLASE; PYRIDOXAL PHOSPHATE.
FT BINDING 242 242 PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ SEQUENCE 413 AA; 47628 MW; 88511B0D CRC32;

Query Match 72.7%; Score 48; DB 1; Length 413;
Best Local Similarity 66.7%; Pred. No. 2.28e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 243 FLCGPM 248
QY 531 YLSCPM 536
|||||

RESULT 12
ID GRAL_MOUSE STANDARD; PRT; 449 AA.
AC Q64018; Q64019;
DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE GLYCINE RECEPTOR ALPHA-1 CHAIN PRECURSOR (48 KD) (STRYCHNINE BINDING
 DE SUBUNIT).
 GN GLRAL.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 95004575.
 RA RYAN S.G., BUCKWALTER M.S., LYNCH J.W., HANDFORD C.A., SEGURA L.,
 RA SHANG R., WASMUTH J.J., CAMPER S.A., SCHOFIELD P., O'CONNELL P.;
 RT "A missense mutation in the gene encoding the alpha 1 subunit of the
 RT inhibitory glycine receptor in the spasmodic mouse.";
 RL NAT. GENET. 7:131-135(1994).
 CC -!- FUNCTION: THE GLYCINE RECEPTOR IS A NEUROTANSMITTER-GATED ION
 CC CHANNEL. BINDING OF GLYCINE TO ITS RECEPTOR INCREASES THE CHLORIDE
 CC CONDUCTANCE AND THUS PRODUCES HYPERPOLARIZATION (INHIBITION OF
 CC NEURONAL FIRING).
 CC -!- SUBUNIT: PENTAMER COMPOSED OF ALPHA (48 KD) AND BETA (58 KD)
 CC SUBUNITS. A THIRD TYPE OF SUBUNIT (93 KD) IS A PERIPHERAL MEMBRANE
 CC PROTEIN ASSOCIATED WITH THE CYTOPLASMIC DOMAINS OF THE RECEPTOR
 CC CORE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- DISEASE: DEFECTS IN GLRAL ARE A CAUSE OF SPASMODIC (SPD), A MOUSE
 CC MUTANT WHICH RESEMBLES TO THE HUMAN NEUROLOGICAL DISEASE,
 CC HYPERKPLEXIA (OR STARTLE DISEASE (STHE)).
 CC -!- THE ALPHA SUBUNIT BINDS STRYCHNINE.
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S73717; G765207; -.
 DR EMBL; S73718; G765209; -.
 DR MGD; MGI-95747; GLRAL.
 DR PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
 DR PFAM; PF00065; neur_chan; 2.
 KW RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
 KW TRANSMEMBRANE; MULTIGENE FAMILY; ALTERNATIVE SPLICING;
 KW DISEASE MUTATION.
 FT SIGNAL 1 28
 FT CHAIN 29 449
 FT DOMAIN 29 247
 FT TRANSMEM 248 274
 FT TRANSMEM 281 298
 FT TRANSMEM 313 336
 FT DOMAIN 337 420
 FT TRANSMEM 421 438
 FT DISULFID 166 180
 FT DISULFID 226 237
 FT CARBOHYD 66 66
 FT BINDING 225 225
 FT BINDING 230 230
 FT VARSPIC 353 353
 SQ SEQUENCE 449 AA; 51725 MW; E5FCD5FE CRC32;
 Query Match 71.2%; Score 47; DB 1; Length 449;
 Best Local Similarity 66.7%; Pred. No. 3.85e+00;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 164 LACPM 169
 I:||||
 Qy 532 LSCPM 537
 RESULT 13

ID GRAL_HUMAN STANDARD; PRT; 449 AA.
 AC P23415;
 DT 01-NOV-1991 (REL. 20, CREATED)
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
 DE GLYCINE RECEPTOR ALPHA-1 CHAIN PRECURSOR (48 KD) (STRYCHNINE BINDING
 DE SUBUNIT).
 GN GLRAL.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 90183975.
 RA GRENNINGLOH G., SCHMIEDEN V., SCHOFIELD P.R., SEEBURG P.H.,
 RA SIDIQUE T., MOHANDAS T.K., BECKER C.M., BETZ H.;
 RT "Alpha subunit variants of the human glycine receptor: primary
 RT structures, functional expression and chromosomal localization of the
 RT corresponding genes.";
 RL EMO J. 9:771-776(1990).
 RN [2]
 RN VARIANTS THE LEU-299 AND GLN-299.
 RX MEDLINE; 94129612.
 RA SHANG R., RYAN S.G., ZHU Y.-Z., HAHN A.F., O'CONNELL P.,
 RA WASMUTH J.J.;
 RT "Mutations in the alpha 1 subunit of the inhibitory glycine receptor
 RT cause the dominant neurologic disorder, hyperekplexia.";
 RL NAT. GENET. 5:351-357(1993).
 RN [3]
 RP VARIANT STHE CYS-307.
 RA SHANG R., RYAN S.G., ZHU Y.-Z., O'CONNELL P., WASMUTH J.J.;
 RT "Mutational and haplotype analysis of the alpha 1 subunit of the
 RT glycine receptor in hyperekplexia patients.";
 RL AM. J. HUM. GENET. 55:A242-A242(1994).
 RN [4]
 RP CHARACTERIZATION OF STHE POSITION 299 VARIANTS.
 RX MEDLINE; 95009923.
 RA LANGOSCH D., LAUBE B., RUNSTROM N., SCHMIEDEN V., BORMANN J.,
 RA BETZ H.;
 RT "Decreased agonist affinity and chloride conductance of mutant
 RT glycine receptors associated with human hereditary hyperekplexia.";
 RL EMO J. 13:4223-4228(1994).
 RN [5]
 RP VARIANT STHE LEU-299.
 RX MEDLINE; 95072599.
 RA SCHORDERET D.F., PESCIA G., BERNASCONI A., REGLI F.;
 RT "An additional family with Startle disease and a Gl192A mutation at
 RT the alpha 1 subunit of the inhibitory glycine receptor gene.";
 RL HUM. MOL. GENET. 3:1201-1201(1994).
 RN [6]
 RP VARIANT STHE ASP-272.
 RX MEDLINE; 95187157.
 RA REES M.I., ANDREW M., JAWAD S., OWEN M.J.;
 RT "Evidence for recessive as well as dominant forms of startle disease
 RT (hyperekplexia) caused by mutations in the alpha 1 subunit of the
 RT inhibitory glycine receptor.";
 RL HUM. MOL. GENET. 3:2175-2179(1994).
 RN [7]
 RP VARIANT STHE CYS-307.
 RX MEDLINE; 95336157.
 RA SHANG R., RYAN S.G., ZHU Y.-Z., FIELDER T.J., ALLEN R.J., FRYER A.,
 RA YAMASHITA S., O'CONNELL P., WASMUTH J.J.;
 RT "Mutational analysis of familial and sporadic hyperekplexia.";
 RL ANN. NEUROL. 38:85-91(1995).
 RN [8]
 RP VARIANT STHE HIS-294.
 RX MEDLINE; 96152139.
 RA MILANI N., DALPRA L., DEL PRETE A., ZANINI R., LARIZZA L.;
 RT "A novel mutation (266 Gln->His) in the alpha 1 subunit of the
 RT inhibitory glycine-receptor gene (GLRAL) in hereditary
 RT hyperekplexia.";
 RL AM. J. HUM. GENET. 58:420-422(1996).
 RN [9]

Query Match 71.2%; Score 47; DB 1; Length 449;
Best Local Similarity 66.7%; Pred. No. 3.85e+00;

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: D00833; D1001166; -
DR EMBL: Y00276; G755778; -
DR EMBL: M63915; G204427; -
DR EMBL: X55246; G56469; -
DR PIR: A27141; A27141;
DR PIR: A38597; A38597;
DR PIR: JN0014; JN0014;
DR PFAM: PS00236; NEUROTR_ION_CHANNEL; 1.
DR PFAM: PF00065; neut_chan; 2.
KW RECEPTOR: POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
KW TRANSMEMBRANE; MULTIGENE FAMILY; ALTERNATIVE SPLICING.
FT SIGNAL 1 28
FT CHAIN 29 449 GLYCINE RECEPTOR ALPHA-1 SUBUNIT.
FT DOMAIN 29 247 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 248 274 PROBABLE.
FT TRANSMEM 281 298 PROBABLE.
FT TRANSMEM 313 336 PROBABLE.
FT DOMAIN 337 420 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 421 438 PROBABLE.
FT DISULFID 166 180 BY SIMILARITY.
FT CARBOHYD 226 237 BY SIMILARITY.
FT BINDING 225 225 STRYCHNINE (OR 230).
FT BINDING 230 230 STRYCHNINE (OR 225).
FT VARSPLIC 353 353 K -> KSPMLNLFQ (IN LONG FORM).
FT CONFLICT 1 22 MYFNTLRFYIWTETIVFFSLAA ->
FT CONFLICT 26 26 MGKSPGLLDYLVWATLFEFP (IN REF. 3).
FT SEQUENCE 449 AA; 51685 MW; 9A517C84 CRC32;

Query Match 71.2%; Score 47; DB 1; Length 449;
Best Local Similarity 66.7%; Pred. No. 3.85e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 164 LACPM 169
QY 532 LSCPM 537

RESULT 15
ID YJB3_YEAST STANDARD; PRT; 515 AA.
AC P47074;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 59.5 KD PROTEIN IN CCT3-CCT8 INTERGENIC REGION.
GN YJL013C OR J1341.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA TO VAN D., PEREA J., JACO C.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: TO YEAST PROTEIN KINASE BUB1 IN ITS NON-CATALYTIC N-TERMINAL DOMAIN

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: Z49288; G1006729; -
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 515 AA; 59521 MW; 032335E9 CRC32;

Query Match 71.2%; Score 47; DB 1; Length 515;
Best Local Similarity 66.7%; Pred. No. 3.85e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 451 YINCPM 456
QY 531 YLSCPM 536

Search completed: Thu Jul 8 19:01:40 1999
Job time : 8 secs.

THIS PAGE BLANK (USPTO)

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:01:58 1999; MasPar time 8.48 Seconds

Tabular output not generated. 51.516 Million cell updates/sec

Title: >US-09-041-236-2
Description: (531-538) from US09041236.pep (27 of 45)
Perfect Score: 66

Sequence: 1 YLSCPMES 8

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertibrate 14:sp.virus

Statistics: Mean 21.833; Variance 26.446; scale 0.826

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	66	100.0	666	4	SEMAPHORIN L.	5.31e-04
2	55	83.3	394	2	HYPOTHETICAL 42.2 KD P	2.29e-01
3	55	83.3	433	2	REGULATORY PROTEIN.	2.29e-01
4	52	78.8	631	5	T12F5.5 PROTEIN.	1.08e+00
5	52	78.8	663	2	CMPC.	1.08e+00
6	51	77.3	356	5	F49D11.1 PROTEIN.	1.79e+00
7	50	75.8	310	8	NADH DEHYDROGENASE SUB	2.96e+00
8	50	75.8	310	8	NADH DEHYDROGENASE SUB	2.96e+00
9	50	75.8	310	8	NADH DEHYDROGENASE SUB	2.96e+00
10	50	75.8	1044	3	NITRITE REDUCTASE.	2.96e+00
11	49	74.2	369	5	COSMID B0222.	4.85e+00
12	49	74.2	379	10	PECTIN METHYLESTERASE	4.85e+00
13	49	74.2	437	5	GAMMA-AMINOBUTYRIC ACI	4.85e+00
14	49	74.2	449	4	GLYCINE RECEPTOR ALPHA	4.85e+00
15	49	74.2	606	5	NUCLEAR HORMONE RECEPT	4.85e+00
16	49	74.2	1119	4	ANKYRIN-LIKE PROTEIN.	4.85e+00
17	48	72.7	211	14	POL POLYPROTEIN (FRAGM	7.91e+00
18	48	72.7	271	5	F35E2.2 PROTEIN.	7.91e+00
19	48	72.7	394	14	POL PROTEIN (FRAGMENT)	7.91e+00
20	48	72.7	394	14	POL PROTEIN (FRAGMENT)	7.91e+00

ID	Query Match	Score	Length	DB	Description	Pred. No.
21	48	72.7	394	14	O56916	7.91e+00
22	48	72.7	394	14	O56912	7.91e+00
23	48	72.7	394	14	O56917	7.91e+00
24	48	72.7	394	14	O56920	7.91e+00
25	48	72.7	394	14	O56913	7.91e+00
26	48	72.7	394	14	O56918	7.91e+00
27	48	72.7	394	14	O56919	7.91e+00
28	48	72.7	478	5	O93933	7.91e+00
29	48	72.7	502	2	O66795	7.91e+00
30	48	72.7	524	5	O18059	7.91e+00
31	48	72.7	560	14	O40444	7.91e+00
32	48	72.7	1037	14	O77373	7.91e+00
33	48	72.7	1033	14	O79666	7.91e+00
34	48	72.7	1039	14	O87108	7.91e+00
35	47	71.2	59	14	O81734	1.28e+01
36	47	71.2	191	5	O17622	1.28e+01
37	47	71.2	348	3	P78791	1.28e+01
38	47	71.2	364	14	O81605	1.28e+01
39	47	71.2	364	14	O81607	1.28e+01
40	47	71.2	364	14	O96848	1.28e+01
41	47	71.2	364	14	O96847	1.28e+01
42	47	71.2	364	14	O81606	1.28e+01
43	47	71.2	445	2	O68141	1.28e+01
44	47	71.2	1449	14	O91080	1.28e+01
45	47	71.2	3015	14	O92532	1.28e+01

ALIGNMENTS

RESULT 1
ID O75326 PRELIMINARY; PRT; 666 AA.
AC O75326;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN L.
GN SEMAL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL: AF030698; G3523115;
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match 100.0%; Score 66; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 5.31e-04;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 563 YLSCPMES 570
QY 531 YLSCPMES 538
|||||||

RESULT 2
ID O86748 PRELIMINARY; PRT; 394 AA.
AC O86748;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 42.2 KD PROTEIN.
GN SC6A9.16.
OS STREPTOMYCES COELICOLOR.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);

```

RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WORLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RC CLARKE K., WOHLDMANN P., ROHLFING T., BAUER C.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RC WATERSTON R.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF039718; G2773235;
SQ SEQUENCE 631 AA; 70685 MW; BE17D0F1 CRC32;

Query Match 78.8%; Score 52; DB 5; Length 631;
Best Local Similarity 62.5%; Pred. No. 1.08e+00;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0

Db 62 YLXCOWDS 69
||| | | |
QY 531 YLSCPMES 538

RESULT 5
ID Q55107 PRELIMINARY; PRT; 663 AA.
AC Q55107;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE CMPC.
GN CMPC.
OS SYNECHOCOCCUS SP. (STRAIN PCC 6301).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC7942.
RA OMATA T., CARLSON T.J., OGAWA T., PIERCE J.;
RT "Sequencing and modification of the gene encoding the 42 kilodalton
RT protein in the cytoplasmic membrane of Synechococcus PCC7942.";
RL PLANT PHYSIOL. 93:305-311(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC7942;
RA OMATA T.;
RT "Characterization of the downstream region of cmpA: identification of
RT a gene cluster encoding a putative permealase of the Cyanobacterium
RT Synechococcus PCC7942.";
RL RES. PHOTOSYN. 3:807-810(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC7942;
RC MEDLEY K.J., MASAMOTO K., SHERMAN D.M., SHERMAN L.A.;
RA REDDY K.J., MASAMOTO K., SHERMAN D.M., SHERMAN L.A.;
RT "DNA sequence and regulation of the gene (cbaA) encoding the
RT 42-kilodalton cytoplasmic membrane carotenoprotein of the
RT Cyanobacterium Synechococcus sp. strain PCC7942.";
RL J. BACTERIOL. 171:3486-3493(1993).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC7942;
RA OMATA T.;
RL SUBMITTED (DEC-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: D26358; D1005930;

```

DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PFAM; PF00005; ABC_tran; 1.
KW ATP-BINDING; TRANSPORT.
SQ SEQUENCE 663 AA; 73492 MW; 5188AB95 CRC32;

Query Match 78.8%; Score 52; DB 2; Length 663;
Best Local Similarity 85.7%; Pred. No. 1.08e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 530 YLSCPKE 536
|||:|:
QY 531 YLSCPME 537

RESULT 6 PRELIMINARY; PRT; 356 AA.

ID O44729;
AC O44729;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F49D11.1 PROTEIN.
GN F49D11.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., FULTON L.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., JEROME L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCHURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA HINDS K., GRAVES T., DUCKELS G., OZERSKY P.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF039711; G2773188; -.
SQ SEQUENCE 356 AA; 41459 MW; 501AF618 CRC32;

Query Match 77.3%; Score 51; DB 5; Length 356;
Best Local Similarity 50.0%; Pred. No. 1.79e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 80 FLSCSMDT 87
:|:|:|:
QY 531 YLSCPME 538

RESULT 7 PRELIMINARY; PRT; 310 AA.

ID O63629;
AC O63629;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE NADH DEHYDROGENASE SUBUNIT 1.
OS DALBULUS GUEVARI.
OC MITOCHONDRION.

OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; HEMIPTERA; EUHEMIPTERA; MEMBRACOIDEA; CICADELLIDAE;
OC DELTOCEPHALINAE; DALBULUS.

RN [1]
RP SEQUENCE FROM N.A.

RA DIETRICH C.H., FITZGERALD S.J., HOLMES J.L., BLACK W.C., NAULT L.R.;
RL ANN. ENTOMOL. SOC. AMER. 0:0-0(1998).

DR EMBL; AF051295; G2961588; -.

DR PROSITE; PS00667; COMPLEX1_NDI_1; 1.

DR PROSITE; PS00668; COMPLEX1_NDI_2; 1.

KW MITOCHONDRION; OXIDOREDUCTASE; NAD; TRANSMEMBRANE.

SQ SEQUENCE 310 AA; 36022 MW; BE2EF0E8 CRC32;

Query Match 75.8%; Score 50; DB 8; Length 310;
Best Local Similarity 83.3%; Pred. No. 2.96e+00;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 71 YLVCPM 76
|||:
QY 531 YLSCPME 536

RESULT 8

ID O63630; PRELIMINARY; PRT; 310 AA.

DT 01-AUG-1998 (TREMBLREL. 07, CREATED)

DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE NADH DEHYDROGENASE SUBUNIT 1.

OS DALBULUS LONGULUS.

OC MITOCHONDRION.

OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;

OC PTERYGOTA; HEMIPTERA; EUHEMIPTERA; MEMBRACOIDEA; CICADELLIDAE;

OC DELTOCEPHALINAE; DALBULUS.

RN [1]
RP SEQUENCE FROM N.A.

RA DIETRICH C.H., FITZGERALD S.J., HOLMES J.L., BLACK W.C., NAULT L.R.;

RL ANN. ENTOMOL. SOC. AMER. 0:0-0(1998).

DR EMBL; AF051296; G2961590; -.

KW MITOCHONDRION.

SQ SEQUENCE 310 AA; 36159 MW; 6FA5D607 CRC32;

Query Match 75.8%; Score 50; DB 8; Length 310;
Best Local Similarity 83.3%; Pred. No. 2.96e+00;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 71 YLVCPM 76
|||:
QY 531 YLSCPME 536

RESULT 9

ID O63626; PRELIMINARY; PRT; 310 AA.

DT 01-AUG-1998 (TREMBLREL. 07, CREATED)

DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE NADH DEHYDROGENASE SUBUNIT 1.

OS DALBULUS ELIMATUS.

OC MITOCHONDRION.

OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;

OC PTERYGOTA; HEMIPTERA; EUHEMIPTERA; MEMBRACOIDEA; CICADELLIDAE;

OC DELTOCEPHALINAE; DALBULUS.

RN [1]
RP SEQUENCE FROM N.A.

RA DIETRICH C.H., FITZGERALD S.J., HOLMES J.L., BLACK W.C., NAULT L.R.;

RL ANN. ENTOMOL. SOC. AMER. 0:0-0(1998).

DR EMBL; AF051292; G2961582; -.

DR PROSITE; PS00667; COMPLEX1_NDI_1; 1.

DR PROSITE; PS00668; COMPLEX1_NDI_2; 1.

KW MITOCHONDRION; OXIDOREDUCTASE; NAD; TRANSMEMBRANE.

SQ SEQUENCE 310 AA; 36327 MW; 083DE482 CRC32;

```
Query Match 75.8%; Score 50; DB 8; Length 310;
Best Local Similarity 83.3%; Pred. No. 2.96e+00;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 71 YLVCPM 76
QY 531 YLSCPM 536

RESULT 10
ID Q00943 PRELIMINARY; PRT; 1044 AA.
AC Q00943; 1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE NITRITE REDUCTASE.
GN YN1.
OS HANSENULA POLYMORPHA (YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; PICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCYC 495;
RA BRITO N., AVILA J., PEREZ M., GONZALEZ C., SIVERTIO J.M.;
RL J. BIOCHEM. 317:89-95(1996).
DR EMBL; Z68122; E211611; -
DR PFAM; PF01077; NIR_SIR; 1.
SQ SEQUENCE 1044 AA; 116574 MW; 59F4D4B1 CRC32;

Query Match 75.8%; Score 50; DB 3; Length 1044;
Best Local Similarity 83.3%; Pred. No. 2.96e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 958 YISCPM 963
QY 531 YLSCPM 536

RESULT 11
ID Q17457 PRELIMINARY; PRT; 369 AA.
AC Q17457; 1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COSMID B0222.
GN B0222.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WORLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA DU Z., GATTUNG S.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U50312; G1226313; -
DR PFAM; PF00086; thyroglobulin_1; 2.
SQ SEQUENCE 369 AA; 41135 MW; 0F5EAFBF CRC32;

Query Match 74.2%; Score 49; DB 5; Length 369;
Best Local Similarity 62.5%; Pred. No. 4.85e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 246 YFSCAMQS 253
QY 531 YLSCPMES 538

RESULT 12
ID Q96497 PRELIMINARY; PRT; 379 AA.
AC Q96497;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE PECTIN METHYLESTERASE (FRAGMENT).
OS SILENE PRATENSIS (WHITE CAMPION) (LYCHNIS ALBA).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC CARYOPHYLLIDAE; CARYOPHYLLALES; CARYOPHYLLACEAE; SILENE.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FLOWERBUD;
RA HINNISDAELS S.J.G.V., NEGRUTIU I.;
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Y08155; E266674; -
DR PFAM; PF01095; Pectinesterase; 1.
DR MENDEL; 11953; MELam.1273; 1.
FT NON_TER 1
SQ SEQUENCE 379 AA; 41878 MW; AB8A3891 CRC32;

Query Match 74.2%; Score 49; DB 10; Length 379;
Best Local Similarity 62.5%; Pred. No. 4.85e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 305 YINCTMES 312
QY 531 YLSCPMES 538

RESULT 13
ID Q08861 PRELIMINARY; PRT; 437 AA.
AC Q08861;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GAMMA-AMINOBUTYRIC ACID-GATED CHLORIDE-ION CHANNEL/RECEPTOR, ZETA
DE SUBUNIT PRECURSOR.
OS LYMAEAE STAGNALIS (GREAT POND SNAIL).
OC EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; BASOMMATOPHORA;
OC LYMAEIDAE; LYMAEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=WHOLE ORGANISM;
RX MEDLINE; 93314766.
RA HUTTON M.L., HARVEY R.J., EARLEY F.G.P., BARNARD E.A., DARLISON M.G.;
RT pattern of gene expression.";
RL FEBS LETT. 326:112-116(1993).
DR EMBL; X71357; G396716; -
DR PFAM; PF00065; neur_chan; 2.
KW VOLTAGE-GATED CHANNEL; RECEPTOR; SIGNAL; TRANSMEMBRANE.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 437 GAMMA-AMINOBUTYRIC ACID-GATED
FT SUBUNIT. CHLORIDE-ION CHANNEL/RECEPTOR, ZETA
```


SQ SEQUENCE 437 AA; 50013 MW; FF9AAD74 CRC32;

Query Match 74.2%; Score 49; DB 5; Length 437;

Best Local Similarity 83.3%; Pred. No. 4.85e+00;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 164 LSCPM 169

|||||

Qy 532 LSCPME 537

RESULT 14

ID O75816 PRELIMINARY; PRT: 449 AA.

AC O75816;

DT 01-NOV-1998 (TREMBREL. 08, CREATED)

DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)

DE GLYCINE RECEPTOR ALPHA 3 SUBUNIT.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;

OC CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 98344067.

RA NIKOLIC Z., LAUBE B., WEBER R.G., LICHTER P., KIOSCHIS P., POUSTKA A.,

RA MULHARDT C., BECKER C.M.;

RT "The Human Glycine Receptor Subunit alpha3 (subtitle: GIRAP3 GENE

RT STRUCTURE, CHROMOSOMAL LOCALIZATION, AND FUNCTIONAL CHARACTERIZATION

RT OF ALTERNATIVE TRANSCRIPTS).";

RL J. BIOL. CHEM. 273:19708-19714(1998).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.

DR EMBL; U93917; G3342236; -

DR PROSITE; PS00236; NEUROTRION_CHANNEL; 1.

KW POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; TRANSMEMBRANE.

SQ SEQUENCE 449 AA; 51963 MW; 3170F74B CRC32;

Query Match 74.2%; Score 49; DB 4; Length 449;

Best Local Similarity 83.3%; Pred. No. 4.85e+00;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 169 LSCPM 174

|||||

Qy 532 LSCPME 537

RESULT 15

ID O77245 PRELIMINARY; PRT: 606 AA.

AC O77245;

DT 01-NOV-1998 (TREMBREL. 08, CREATED)

DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)

DE NUCLEAR HORMONE RECEPTOR E75.

GN E75.

OS METAPENAEUS ENSIS.

OC EUKARYOTA; METAZOA; ARTHROPODA; CRUSTACEA; MALACOSTRACA;

OC EUMALACOSTRACA; EUCARIDA; DECAPODA; DENDROBRANCHIATA; PENAEIDAE;

OC METAPENAEUS.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-PLEPOD (SWIMMING LEG);

RA CHAN S.-M.;

RT "Cloning of a shrimp (Metapenaeus ensis) cDNA encoding a nuclear

RT receptor superfamily member: An insect homologue of E75 gene.;"

RL FEBS LETT. 0:0-0(1998).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

DR EMBL; AF092946; G3694976; -

DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.

KW RECEPTOR; TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN;

KW ZINC-FINGER.

SQ SEQUENCE 606 AA; 67689 MW; 05E98AC3 CRC32;

Query Match

74.2%; Score 49; DB 5; Length 606;

Best Local Similarity 50.0%; Pred. No. 4.85e+00;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 180 YLACPLNP 187

|||||

Qy 531 YLSCPME 538

Search completed: Thu Jul 8 19:02:11 1999

Job time : 13 secs.

THIS PAGE BLANK (USPTO)

•

WQSRFH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:05:12 1999; MasPar time 7.37 Seconds
Tabular output not generated. 25.966 Million cell updates/sec

Title: >US-09-041-236-2
Description: (554-562) from US09041236.pep (28 of 45)
Perfect Score: 66
Sequence: 1 SCEPGHQSP 9
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 15.079; Variance 43.046; scale 0.350

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	47	71.2	421	7 R33733	Rat intrinsic factor.	6.92e+01
2	47	71.2	557	5 R26323	B6GP.	6.92e+01
3	46	69.7	788	5 R26322	B6H.	9.02e+01
4	45	68.2	95	22 W15400	OPNPV Type 145 ORF in	1.17e+02
5	45	68.2	510	7 R34674	Insulinoma-associated	1.17e+02
6	45	68.2	538	23 W10026	Partial sporozoite at	1.17e+02
7	44	66.7	150	6 R13174	Human bone morphogeni	1.52e+02
8	44	66.7	977	8 R51282	Helminth aminopeptida	1.52e+02
9	44	66.7	2189	1 R05222	Antigen GX5401FL enco	1.52e+02
10	43	65.2	20	38 W67528	IGF-1/IGFBP inhibitor	1.97e+02
11	43	65.2	21	22 W05796	Presenilin-2 residues	1.97e+02
12	43	65.2	414	22 W05766	Presenilin-2 delta263	1.97e+02
13	43	65.2	447	24 W28508	Full AD4/AD3LP sequen	1.97e+02
14	43	65.2	448	30 W23967	Human presenilin-2.	1.97e+02
15	43	65.2	448	24 W11321	Human AD4 protein.	1.97e+02
16	43	65.2	448	22 W05764	Presenilin-2 N41I mu	1.97e+02

17	43	65.2	448	22 W05763	Presenilin-2 M239V mu	1.97e+02
18	43	65.2	448	22 W05765	Presenilin-2 I420T mu	1.97e+02
19	43	65.2	448	22 W05762	Human presenilin-2.	1.97e+02
20	42	63.6	546	22 W40115	Human alpha-5(IV) col	2.55e+02
21	41	62.1	206	32 W48977	Mouse OX40 extracellu	3.29e+02
22	41	62.1	206	16 R81881	Mouse type-II membran	3.29e+02
23	41	62.1	232	31 W37841	Human 312C2 protein f	3.29e+02
24	41	62.1	277	14 R74737	ACT-4 cell surface re	3.29e+02
25	41	62.1	277	14 R76996	Deduced sequence enco	3.29e+02
26	41	62.1	277	15 R79904	ACT-4-h-1 receptor se	3.29e+02
27	41	62.1	383	10 R56166	Neuroendocrine tumor	3.29e+02
28	41	62.1	438	32 W48976	OX40/Fc mutein.	3.29e+02
29	41	62.1	438	16 R81882	Plasmid pDC406/OX40/F	3.29e+02
30	41	62.1	610	24 W18839	E-selectin.	3.29e+02
31	41	62.1	610	1 R05538	Endothelial leukocyte	3.29e+02
32	41	62.1	610	2 R08116	Endothelial cell-leuc	3.29e+02
33	41	62.1	610	29 W46733	Amino acid sequence o	3.29e+02
34	41	62.1	610	1 R05494	Endothelial leukocyte	3.29e+02
35	41	62.1	1342	3 R13833	HSP-3 epithelial grow	3.29e+02
36	41	62.1	1342	16 R88453	erbB-3 polypeptide.	3.29e+02
37	41	62.1	1342	36 W69406	erbB-3 glycoprotein c	3.29e+02
38	41	62.1	1343	3 R12608	EGFR-R erbB-3 clone E	3.29e+02
39	41	62.1	1833	14 R79478	Mouse LTRP-2.	3.29e+02
40	40	60.6	224	15 R82989	P815A antigen precurs	4.24e+02
41	40	60.6	224	12 R74004	Tumour rejection anti	4.24e+02
42	40	60.6	730	1 P80618	Human Bone Morphogeni	4.24e+02
43	40	60.6	730	22 W13669	C-proteinase encoded	4.24e+02
44	40	60.6	1212	29 W44299	Human serrate 2.	4.24e+02
45	40	60.6	1358	24 W18824	Human restrictin.	4.24e+02

ALIGNMENTS

RESULT 1
ID R33733 standard; Protein; 421 AA.
AC R33733;
DT 23-JUL-1993 (first entry)
DE Rat intrinsic factor.
KW Human; intrinsic factor; probe; clone; rat.
OS Rattus norvegicus.
FH Key Location/Qualifiers
FT peptide 1..22 /note= "signal peptide"
FT protein 23..421 /note= "Mature intrinsic factor"
FT J05049478-A.
PD 02-MAR-1993.
PF 06-JUN-1991; 134788.
PR 06-JUN-1991; JP-134788.
PA (KANF) KANEKA CORP.
DR WPI; 93-111881/14.
PT DNA sequence coding human intrinsic factor - is used for cloning
PT human intrinsic factor DNA using the new DNA as a probe
PS Disclosure; Page 9-10; 11pp; Japanese.
SC This sequence represents rat intrinsic factor. This sequence was
CC isolated using probes based on the human intrinsic factor cDNA.
SQ Sequence 421 AA;

Query Match 71.2%; Score 47; DB 7; Length 421;
Best Local Similarity 55.8%; Pred. No. 6.92e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 291 tcepdhevp 299
QY 554 SCEPGHQSP 562

RESULT 2
ID R26323 standard; Protein; 557 AA.
AC R26323;
DT 29-JAN-1993 (first entry)
DE B6GP.
KW Beta-6; beta-1; beta-2; beta-3; cytoplasmic tail; ligand; vitronectin;

KW epithelial-derived; tumour cells; fibronectin.
 OS Cavia porcellus.
 PN WO9212236-A.
 PD 23-JUL-1992.
 PF 11-JAN-1991; U00236.
 PR 11-JAN-1991; WO-002362.
 PA (SCRI) SCRIPPS CLINIC & RES CENT.
 PA (REGC) UNIV CALIFORNIA.
 PI Pytela R, Quaranta V, Sheppard D;
 DR WPI: 92-284332/34.
 DR N-PSDB: Q27643.
 PT New integrin beta sub-unit and its nucleic acid - forms
 PT hetero-dimers with sub-units alpha-V and alpha-E, useful as a
 PT diagnostic

PS Disclosure: Fig 3: 43pp; English.
 CC The sequences given in R26322-23 are human and guinea pig integrin beta
 CC subunit, beta-6 proteins. The beta-6 subunit is a surface receptor
 CC which is useful in mediating critical aspects of cell processes in
 CC conjunction with an integrin alpha subunit. Beta-6 is clearly
 CC different from beta-1, beta-2, beta-3 and other beta subunits which
 CC have been discovered. The 11 amino acid carboxyl-terminal extension
 CC distinguishes it from other beta subunits. This cytoplasmic tail
 CC indicates that beta-6 regulates signal transduction differently from
 CC other beta subunits. Beta-6 has been isolated from epithelial-derived
 CC tumour cells in association with the integrin alpha subunit, alpha-V.
 CC Beta-6 can associate with a variety of alpha subunits to form a
 CC functional integrin. Alpha-V/beta-6 can bind extracellular matrix
 CC molecules, eg. ligands, pref. containing the sequence Arg-Gly-Asp,
 CC such as vitronectin or fibronectin.
 SQ Sequence 557 AA;

Query Match 71.2%; Score 47; DB 5; Length 557;
 Best Local Similarity 55.6%; Pred. No. 6.92e+01;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 318 acnpgmhgp 326
 :|||:|
 QY 554 SCEPGHQP 562

RESULT 3

ID R26322 standard; Protein: 788 AA.
 AC R26322;
 DT 29-JAN-1993 (first entry)
 DE B6H.
 KW Beta-6; beta-1; beta-2; beta-3; cytoplasmic tail; ligand; vitronectin;
 KW epithelial-derived; tumour cells; fibronectin.
 OS Homo sapiens.
 PN WO9212236-A.
 PD 23-JUL-1992.
 PF 11-JAN-1991; U00236.
 PR 11-JAN-1991; WO-002362.
 PA (SCRI) SCRIPPS CLINIC & RES CENT.
 PA (REGC) UNIV CALIFORNIA.
 PI Pytela R, Quaranta V, Sheppard D;
 DR WPI: 92-284332/34.
 DR N-PSDB: Q27642.
 PT New integrin beta sub-unit and its nucleic acid - forms
 PT hetero-dimers with sub-units alpha-V and alpha-E, useful as a
 PT diagnostic
 PS Claim 2; Fig 3: 43pp; English.
 CC The sequences given in R26322-23 are human and guinea pig integrin beta
 CC subunit, beta-6 proteins. The beta-6 subunit is a surface receptor
 CC which is useful in mediating critical aspects of cell processes in
 CC conjunction with an integrin alpha subunit. Beta-6 is clearly
 CC different from beta-1, beta-2, beta-3 and other beta subunits which
 CC have been discovered. The 11 amino acid carboxyl-terminal extension
 CC distinguishes it from other beta subunits. This cytoplasmic tail
 CC indicates that beta-6 regulates signal transduction differently from
 CC other beta subunits. Beta-6 has been isolated from epithelial-derived
 CC tumour cells in association with the integrin alpha subunit, alpha-V.
 CC Beta-6 can associate with a variety of alpha subunits to form a
 CC functional integrin. Alpha-V/beta-6 can bind extracellular matrix

CC molecules, eg. ligands, pref. containing the sequence Arg-Gly-Asp,
 CC such as vitronectin or fibronectin.
 SQ Sequence 788 AA;

Query Match 69.7%; Score 46; DB 5; Length 788;
 Best Local Similarity 55.6%; Pred. No. 9.02e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 480 achpgmhgp 488
 :|||||:
 QY 554 SCEPGHQP 562

RESULT 4

ID W15400 standard; Protein: 95 AA.
 AC W15400;
 DT 11-JUL-1997 (first entry)
 DE OpNPV Type 145 ORF insecticidal polypeptide.
 KW Poxvirus; entomopoxvirus; EPV; nuclear polyhedrosis virus; OpNPV;
 KW baculovirus; insecticide; transgenic plant; insect resistance.
 OS Orgyia pseudotsugata nuclear polyhedrosis virus.
 FH Key Location/Qualifiers
 FT misc_difference 32
 FT /label= conserved Cys residue
 FT misc_difference 38
 FT /label= conserved Gly residue
 FT misc_difference 43
 FT /label= conserved Pro residue
 FT misc_difference 46
 FT /label= conserved Cys residue
 FT misc_difference 50
 FT /label= conserved Tyr residue
 FT misc_difference 52
 FT /label= conserved Cys residue
 FT misc_difference 60
 FT /label= conserved Cys residue
 FT misc_difference 65
 FT /label= conserved Glu residue
 FT misc_difference 73
 FT /label= conserved Cys residue
 FT misc_difference 76
 FT /label= conserved Ile residue
 FT misc_difference 84
 FT /label= conserved Gly residue
 FT misc_difference 85
 FT /label= conserved Cys residue
 FT misc_difference 87
 FT /label= conserved Ala residue
 FT /label= conserved Ala residue
 PN WO9712906-A1.
 PD 10-APR-1997.
 PE 27-SEP-1996; AU0615
 PR 29-SEP-1995; AU-005702.
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PI Dali DJ;
 DR WPI: 97-226160/20.
 PT New Heliothis armigera entomopoxvirus 11 kD polypeptide - useful in
 PT insecticidal compositions for controlling insect pests
 PS Claim 18; Fig 1b; 31pp; English.
 CC A polypeptide sequence (W15400) is deduced from the Type 145 open
 CC reading frame (ORF) of Orgyia pseudotsugata nuclear polyhedrosis
 CC virus (OpNPV). It shows homology to an 11 kDa polypeptide (see also
 CC W15397) from Heliothis armigera entomopoxvirus and to deduced amino
 CC acid sequences from OpNPV ORF 150 and Autographa californica and
 CC Bombyx mori NPV type 145 and 150 ORFs (see also W15398-99,
 CC W15401-402). A generic sequence (W15396) is provided for this
 CC newly recognised family of insecticidal polypeptides. The
 CC polypeptides, or their insecticidally-active fragments, can be
 CC expressed in host cells, esp. Sf9 or H2-BCIRL-AM1 insect cells, for
 CC use in controlling the proliferation of insect pests. They can also
 CC be expressed in transgenic plants to improve insect resistance.
 SQ Sequence 95 AA;

Query Match 68.2%; Score 45; DB 22; Length 95;

```
Best Local Similarity 66.7%; Pred. No. 1.17e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 60 cdpghe 65
Oy 555 CEPGHQ 560
I:||||

RESULT 5
ID R34674 standard; Protein: 510 AA.
AC R34674;
DT 12-AUG-1993 (first entry)
DE Insulinoma-associated antigen (IA-1).
KW Neuroendocrine; zinc-finger; DNA binding; pro-hormone;
KW regulatory factor; islet; cell transformation; antibody;
KW probe; hybridisation; ELISA; Western Blot; diagnosis;
KW insulinoma; tumour.
OS Homo sapiens.
FH Key Location/Qualifiers
FT domain 1..250
FT domain /label= N-terminal
FT domain 251..510
FT domain /label= C-terminal
PN US7901715-A.
PD 01-APR-1993.
PF 17-JUN-1992; 901715.
PR 17-JUN-1992; US-901715.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Lan M, Notkins A;
DR WPI; 93-152041/18.
DR N-PSDB; Q41087.
PT Recombinant polynucleotide having sequence derived from
PT insulinoma-associated antigen - useful for diagnosis and
PT identification of insulinoma and neuroendocrine tumour
PT Disclosure; Page 34-36; 46pp; English.
PS The IA-1 gene is uniquely expressed in malignant neuroendocrine
CC tissues and/or cell lines. Its sequence includes zinc-finger DNA-
CC binding motifs and dibasic amino acid pro-hormone conversion sites.
CC The restricted tissue distribution and unique sequence motifs of IA-1
CC suggest that IA-1 may function as a regulatory factor in islet
CC cell transformation. The gene, its translation prod. and antibodies
CC to this can be used as probes (either by hybridisation or in ELISA
CC or Western Blot assays) to diagnose correctly human insulinomas and
CC neuroendocrine tumours. The antibodies to IA-1 can be used to
CC substantially purify or separate IA-1 from surrounding native
CC protein.
SQ Sequence 510 AA;

Query Match 68.2%; Score 45; DB 7; Length 510;
Best Local Similarity 55.6%; Pred. No. 1.17e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 71 acapgpqpp 79
Oy 554 SCEPGHQSP 562
I:|||||

RESULT 6
ID W10026 standard; Protein: 538 AA.
AC W10026;
DT 16-SEP-1997 (first entry)
DE Partial sporozoite attachment protein 1.
KW SAP-1; pig; vaccine; infection; antigen; epitope; repeat motif;
KW coccidiosis; merozoite.
OS Isospora suis.
FH Key Location/Qualifiers
FT peptide 288..314
FT peptide /label= repeat_I_motif
FT peptide 315..341
FT peptide /label= repeat_I_motif
FT peptide 342..368
FT peptide /label= repeat_I_motif
FT peptide 369..395

Best Local Similarity 68.2%; Score 45; DB 23; Length 538;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 6 gcnbgdesp 14
Oy 554 SCEPGHQSP 562
I:|||||

RESULT 7
ID R31374 standard; Protein: 150 AA.
AC R31374;
DT 24-JUN-1993 (first entry)
DE Human bone morphogenic protein-9 (BMP-9).
KW bone formation; cartilage formation; wound healing; tissue repair;
KW surgery; fracture treatment; periodontal disease; osteoporosis;
KW increase neuronal survival; transplantation; nerves; nervous system.
OS Homo sapiens.
PN W09300432-A.
PD 07-JAN-1993.
PF 25-JUN-1992; U05374.
PR 25-JUN-1991; US-720590.
PA (GEMV ) GENETICS INST. INC.
PI Celeste AJ, Wozney JM;
DR WPI; 93-036379/04.
DR N-PSDB; Q35244.
PT New BMP-9 polypeptide(s) induce bone and cartilage formation -
PT used for treating osteoporosis and fractures, healing wounds and
PT increasing neuronal survival
PS Example 2; Fig 3; 60pp; English.
CC This sequence is human bone morphogenic protein-9. The protein can
CC be used to induce bone and/or cartilage formation and in wound healing
CC and tissue repair. It can be used in surgery of the treatment of
CC fractures, periodontal disease or osteoporosis. The protein can also
CC increase neuronal survival and can be used in transplantation and
CC treatment of conditions exhibiting a decrease in neuronal survival.
SQ Sequence 150 AA;
```

Query Match 66.7%; Score 44; DB 6; Length 150;
 Best Local Similarity 75.0%; Pred. No. 1.52e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 114 ceptkqsp 121
 ||| |||
 QY 555 CEPGHQSP 562

RESULT 8
 ID R51282 standard; Protein; 977 AA.
 AC R51282;
 DT 31-MAY-1994 (first entry)
 DE Helminth aminopeptidase H110D variant encoded by the H11-1 gene.
 KW Helminth; aminopeptidase; H11-3; H11-2; H11-1; integral membrane;
 KW alpha-amino acyl peptide hydrolase (microsomal); multigene family;
 KW antigen; vaccine; parasite; human; virus; microbe.
 OS Haemonchus contortus.
 PN W09323542-A.
 PD 25-NOV-1993.
 PF 07-MAY-1993; G00943.
 PR 08-MAY-1992; GB-009993.
 PA (AGRI-) AGRIC & FOOD RES COUNCIL.
 PI Graham M, Knox DP, Munn EA, Newton SE, Oliver JJ;
 PI Smith TS;
 DR WPI: 93-386574/48.
 DR N-PSDB: Q52491.
 PT New DNA encoding amino-peptidase from Haemonchus contortus and
 PT derived peptide(s) - useful in protective vaccines, for new
 PT vectors, transformed cells and oligosaccharide, for incorporation
 PT in virus or microbe
 PS Claim 6; Page 78-80; 137pp; English.
 CC The sequences given in R51280-82 represent variants of the helminth
 CC aminopeptidase H110D encoded by the genes H11-3, -2 and -1 respectively.
 CC H110D is a protein doublet which shows homology to a family of integral
 CC membrane aminopeptidases. The differences between the coding sequences
 CC and the translated products, can be attributed to different mRNAs of
 CC the multigene family, and also to different variants of the H110D-
 CC encoding sequence being present at different stages of the life
 CC cycle, or in strains differing in geographical origin. Antigenic
 CC fragments of these aminopeptidases may be used in vaccines to stimulate
 CC immune response against helminth parasites in humans or other animals.
 CC The DNA sequences encoding these proteins may be incorporated into a
 CC virus or microbe and used in a similar manner.
 SQ Sequence 977 AA;

Query Match 66.7%; Score 44; DB 8; Length 977;
 Best Local Similarity 57.1%; Pred. No. 1.52e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 787 cqpqgqa 793
 |::||:
 QY 555 CEPGHQS 561

RESULT 9
 ID R05222 standard; protein; 2189 AA.
 AC R05222;
 DT 02-AUG-1990 (first entry)
 DE Antigen GX5401FL encoded by Eimeria tenella genomic DNA
 KW Eimeria tenella; antigen GX5401FL; antigen GX5401; avian coccidiosis.
 OS Eimeria tenella.
 PN W09000403-A.
 PD 25-JAN-1990.
 PF 05-JUL-1989; U02918.
 PR 05-JUL-1988; US-215162.
 PA (GENE-) Genex Corp.
 PI Anderson DM, McCandless RJ, Strausberg SL, Strausberg RL;
 DR WPI: 90-051586/07.
 DR N-PSDB: Q03324.
 PT Cloned gene or fragment encoding antigenic protein -
 PT which binds with antibodies against avian coccidia, and

PT transformed cells used in vaccine
 PS Claim 10; Page 93; Fig 14; 134pp; English.
 CC It is encoded by an open reading frame contained within the sequence of
 CC clone 533 which was derived from an E. tenella genomic library screened
 CC with radioactively labelled cDNA encoding the GX5401 antigen. It is of
 CC about 250 kd. It carries several repeated peptide sequences and
 CC is rich in cysteine residues. The open reading frame also encodes a
 CC potential signal sequence for protein secretion. Also new are an
 CC expression vector contg. cloned gene, and host cells transformed with
 CC the vector. The transformed cells are used in a vaccine to immunise
 CC birds against avian coccidiosis. By labelling the peptides, they can be
 CC used as a type-specific probe. May also be used in an assay to detect
 CC Ab against the coccidia. The Abs are used to identify transformed cells
 CC contg. the DNA.
 SQ Sequence 2189 AA;

Query Match 66.7%; Score 44; DB 1; Length 2189;
 Best Local Similarity 44.4%; Pred. No. 1.52e+02;
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 371 tcnpgyeap 379
 |::||::|
 QY 554 SCEPGHQP 562

RESULT 10
 ID W67528 standard; peptide; 20 AA.
 AC W67528;
 DT 02-MAR-1999 (first entry)
 DE IGF-1/IGFBP inhibitory peptide 23D3.6.
 KW Inhibition; interaction; insulin-like growth factor; IGF; receptor;
 KW binding protein; serum; tissue; insulin; plasma; growth hormone;
 KW glucose; secretion; blood; hyperglycaemia; obesity.
 OS Synthetic.
 PN W09845427-A2.
 PD 15-OCT-1998.
 PF 31-MAR-1998; U06514.
 PR 04-APR-1997; US-825852.
 PA (GENE-) GENENTECH INC.
 PI Clark RG, Lowman HB, Robinson ICAP;
 PI WPI: 98-583196/49.
 PT Inhibitors of interaction between insulin-like growth factor -
 PT useful for, e.g. treating or preventing hyperglycaemia, obesity and
 PT neurological disorders, and are optionally formulated with
 PT thiazolidinone
 PS Example 7; Page 57; 133pp; English.
 CC The invention relates to the isolation of compounds, especially peptides,
 CC that: (a) inhibit interaction between an insulin-like growth factor (IGF)
 CC with any one of its binding proteins (IGFBP), and (b) do not bind to a
 CC human IGF receptor (hIGFR), e.g. see W67476-W67491 and W67503-W67565.
 CC The peptides are produced synthetically or are isolated from peptide
 CC libraries where the encoding sequence is generated so that the resulting
 CC peptide produced retains its structure in solution. This sequence
 CC represents a peptide isolated from a g8 display peptide library which
 CC binds the IGF-1 binding protein IGFBP-3. The compounds are used to
 CC increase serum and tissue levels of active IGF-1, preferably also to
 CC reduce insulin and plasma growth hormone (GH) secretion, and blood
 CC glucose levels, without direct stimulation of secretion or release of
 CC endogenous GH. Particularly the compounds are used to treat or prevent
 CC hyperglycaemia, obesity-related disorders, neurological, cardiac,
 CC anabolic, renal and immunological diseases.
 SQ Sequence 20 AA;

Query Match 65.2%; Score 43; DB 38; Length 20;
 Best Local Similarity 83.3%; Pred. No. 1.97e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 7 cesghq 12
 ||::|||
 QY 555 CEPGHQ 560

RESULT 11

ID W05796 standard; peptide: 21 AA.
AC W05796;
DT 28-JUL-1997 (first entry)
DE Presenilin-2 residues 23-45.
KW Presenilin-2; human; hPS1-1; hPS1-2; PS-2; integral membrane protein; AD;
KW familial Alzheimer's disease; cerebral haemorrhage; schizophrenia;
KW depression; antibody; gene expression modulator; therapy; muten.
OS Homo sapiens.
PN W0634099-A2.
PD 31-OCT-1996.
PF 29-APR-1996; CA0263.
PR 28-APR-1995; US-431048.
PR 28-JUN-1995; US-496841.
PR 31-JUL-1995; US-509359.
PA (HSCR-) HSC RES & DEV LP.
PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.
PI Fraser PE, Rommens JM, St George-Hyslop PH;
DR WPI; 96-497631/49.
PT New presenilin genes - useful for diagnosis, therapy and drug
PT screening of familial Alzheimer's disease, cerebral disorders, etc.
PS Claim 72; Page -; 178pp; English.
CC W05796-W05808 represent antigenic fragments of the human presenilin-2
CC protein (see W05762 for wild type sequence). The presenilins are a family
CC of highly conserved integral membrane proteins with a common structural
CC motif, common alternate splicing patterns, and common mutational hot spot
CC regions. Mutations in PS genes are implicated in familial Alzheimer's
CC disease (AD) and possibly other diseases such as cerebral haemorrhage,
CC schizophrenia, depression etc., so detection of mutations in the DNA
CC encoding the wild type sequences can be used for diagnosis of these
CC diseases. The wild type proteins, or vectors that express them or
CC containing antisense sequences, antibodies selective for these mutant
CC forms of the proteins and modulators of PS gene expression are
CC potentially useful for treatment of AD etc. Transgenic animals are useful
CC as models for drug screening. The antibodies can also be used e.g. for
CC affinity purification and in immunoassays.
SQ Sequence 21 AA;

Query Match 65.2%; Score 43; DB 22; Length 21;
Best Local Similarity 55.6%; Pred. NO. 1.97e-02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 6 scqegrqgp 14
||:|:|:
QY 554 SCEPGHQP 562

RESULT 12
ID W05766 standard; Protein: 414 AA.
AC W05766;
DT 25-JUL-1997 (first entry)
DE Presenilin-2 delta263-296 mutation.
KW Presenilin-2; human; hPS1-1; hPS1-2; PS-2; integral membrane protein; AD;
KW familial Alzheimer's disease; cerebral haemorrhage; schizophrenia;
KW depression; antibody; gene expression modulator; therapy; muten.
OS Homo sapiens.
PN W0634099-A2.
PD 31-OCT-1996.
PF 29-APR-1996; CA0263.
PR 28-APR-1995; US-431048.
PR 28-JUN-1995; US-496841.
PR 31-JUL-1995; US-509359.
PA (HSCR-) HSC RES & DEV LP.
PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.
PI Fraser PE, Rommens JM, St George-Hyslop PH;
DR WPI; 96-497631/49.
PT New presenilin genes - useful for diagnosis, therapy and drug
PT screening of familial Alzheimer's disease, cerebral disorders, etc.
PS Claim 4; Page -; 178pp; English.
CC W05763-W05766 represent mutated versions of the human presenilin-2
CC protein (see W05762 for wild type sequence). The presenilins are a family

CC of highly conserved integral membrane proteins with a common structural
CC motif, common alternate splicing patterns, and common mutational hot spot
CC regions. Mutations in PS genes are implicated in familial Alzheimer's
CC disease (AD) and possibly other diseases such as cerebral haemorrhage,
CC schizophrenia, depression etc., so detection of mutations in the DNA
CC encoding the wild type sequences can be used for diagnosis of these
CC diseases. The wild type proteins, or vectors that express them or
CC containing antisense sequences, antibodies selective for these mutant
CC forms of the proteins and modulators of PS gene expression are
CC potentially useful for treatment of AD etc. Transgenic animals are useful
CC as models for drug screening. The antibodies can also be used e.g. for
CC affinity purification and in immunoassays.
SQ Sequence 414 AA;

Query Match 65.2%; Score 43; DB 22; Length 414;
Best Local Similarity 55.6%; Pred. NO. 1.97e-02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 30 scqegrqgp 38
||:|:|:
QY 554 SCEPGHQP 562

RESULT 13
ID W28508 standard; Protein: 447 AA.
AC W28508;
DT 07-DEC-1997 (first entry)
DE Full AD4/AD3LP sequence.
KW AD3; AD4/AD3LP; Alzheimer's disease; chromosome; missegregation;
KW presenilin; inhibitor; AD; trisomy 21; ss.
OS Homo sapiens.
PN W09707213-A2.
PD 27-FEB-1997.
PF 15-AUG-1996; U13314.
PR 16-AUG-1995; US-002448.
PA (HARD) HARVARD COLLEGE.
PI Li J, Potter H;
DR WPI; 97-165297/15.
DR N-PSDB; T87426.
PT Identifying genes which cause chromosome missegregation - useful for
PT identifying causes of and treatments for diseases, e.g. Alzheimer's
PT disease, cancer and ageing
PS Claim 29; Fig 29; 77pp; English.
CC Identifying genes which cause improper chromosome segregation,
CC screening for inhibitors of chromosome missegregation and processes
CC caused by genes encoding chromosome missegregation promoters
CC was exemplified using Alzheimer's disease. The sequences
CC given in T87401 to T87426 can be used in the above methods.
CC It is not clear from the figure legend, the figure and the
CC disclosure of the specification which sequence of Fig 1 and Fig 28
CC is the AD4/AD3LP or the AD3 sequence.
SQ Sequence 447 AA;

Query Match 65.2%; Score 43; DB 24; Length 447;
Best Local Similarity 55.6%; Pred. NO. 1.97e-02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 30 scqegrqgp 38
||:|:|:
QY 554 SCEPGHQP 562

RESULT 14
ID W23967 standard; Protein: 448 AA.
AC W23967;
DT 20-JUL-1998 (first entry)
DE Human presenilin-2.
KW Presenilin-1; PS1 gene; human; familial Alzheimer's disease; FAD;
KW cerebral haemorrhage; schizophrenia; depression; epilepsy;
KW mental retardation; diagnosis; therapy; transgenic animal.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Domain 88..106

FT FT /label= TM1
FT FT /note= "transmembrane domain 1"
FT FT 107..134
FT FT /label= TM1-2
FT FT /note= "hydrophilic loop"
FT FT 135..160
FT FT /label= TM2
FT FT /note= "transmembrane domain 2"
FT FT 161..169
FT FT /label= TM2-3
FT FT /note= "hydrophilic loop"
FT FT 170..189
FT FT /label= TM3
FT FT /note= "transmembrane domain 3"
FT FT 190..200
FT FT /label= TM3-4
FT FT /note= "hydrophilic loop"
FT FT 201..218
FT FT /label= TM4
FT FT /note= "transmembrane domain 4"
FT FT 219..224
FT FT /label= TM4-5
FT FT /note= "hydrophilic loop"
FT FT 225..244
FT FT /label= TM5
FT FT /note= "transmembrane domain 5"
FT FT 245..249
FT FT /label= TM5-6
FT FT /note= "hydrophilic loop"
FT FT 250..268
FT FT /label= TM6
FT FT /note= "transmembrane domain 6"
FT FT 269..387
FT FT /label= TM6-7
FT FT /note= "hydrophilic loop"
FT FT 388..409
FT FT /label= TM8
FT FT /note= "transmembrane domain 8"
FT FT Misc_difference 141
FT FT /note= "Asn141lle mutation site (Claim 19)"
FT FT Misc_difference 239
FT FT /note= "Met239Val mutation site (Claim 19)"
FT FT Misc_difference 420
FT FT /note= "Ile420Thr mutation site"
PN WO9801549-A2.
PD 15-JAN-1998.
PF 04-JUL-1997; CA0475.
PR 02-JAN-1997; US-034590.
PR 05-JUL-1996; US-021673.
PR 12-JUL-1996; US-021700.
PR 08-NOV-1996; US-029895.
PA (HSCR-) HSC RES & DEV LP.
PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.
PI Fraser PE, Rommens JM, St George-Hyslop PH;
DR WPI: 98-286355/25.
DR N-PSDB: V04669.
PT New isolated mutant presenilin-1 genes - useful for developing
PT products for use in detection, diagnosis and therapy of Alzheimer's
PT disease and for drug screening
PS Claim 19; Page 203-204; 238pp; English.
CC This polypeptide comprises human presenilin-2 (hps2). Its amino
CC acid sequence was deduced from an isolated cDNA clone (see V04669).
CC Human and murine presenilin-1 sequences are also provided (see
CC W23964-66). Mutations in the PS-1 and PS-2 genes are linked to
CC the development in humans of forms of familial Alzheimer's disease
CC (FAD) and may be causative of other disorders, e.g. cognitive,
CC intellectual, neurological or physiological disorders such as
CC cerebral haemorrhage, schizophrenia, depression, mental retardation
CC and epilepsy. Use of the nucleic acids and proteins comprising or
CC derived from the presenilins is made in screening and diagnosing
CC FAD, identifying and developing therapeutics for treatment of FAD,
CC and in producing cell lines and transgenic animals useful as models
CC of FAD. Methods for identifying substances that bind to, or

CC modulate the activity of a presenilin protein, and methods for
CC identifying substances that affect the interaction of a
CC presenilin-interacting protein with a presenilin protein are also
CC disclosed. 448 AA;
SQ Sequence

Query Match 65.2%; Score 43; DB 30; Length 448;
Best Local Similarity 55.6%; Pred. No. 1.97e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 30 scqegrggp 38
||: |::|
QY 554 SCEPGHQSP 562

RESULT 15
ID W11321 standard; Protein; 448 AA.
AC W11321;
DT 10-NOV-1997 (first entry)
DE Human AD4 protein.
KW Autosomal dominant early-onset Alzheimer's Disease; AD4; STM2;
KW neurodegeneration; senile dementia; human chromosome 1;
KW Volga German kindred; VG.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 1..448
FT /label= membrane_protein
FT /note= "Sequence contains 7 hydrophobic segments
FT presumed to be membrane-spanning regions"
FT Misc_difference 141
FT /note= "Wild-type Asn is mutated to Ile in VG
FT individuals"
FT 222..225
FT /label= conserved_tetramer
FT /note= "Conserved in the human and mouse AD3 genes"
FT 271..274
FT /label= conserved_tetramer
FT /note= "Conserved in the human and mouse AD3 genes"
FT 386
FT modified_site
FT /label= N-glycosylation_site
FT /note= "putative"
PN WO9703192-A2.
PD 30-JAN-1997.
PF 05-JUL-1996; U11386.
PR 14-AUG-1995; US-002328.
PR 07-JUL-1995; US-000956.
PR 28-JUL-1995; US-001675.
PR 11-AUG-1995; US-002174.
PA (DARW-) DARWIN MOLECULAR CORP.
PA (GEO) GEN HOSPITAL CORP.
PA (VAME-) VA MEDICAL CENT.
PI Bird TD, Galas DJ, Levy-Lahad E, Mulligan J, Schellenberg GD;
PI Tanzi RE, Wasco W;
DR WPI: 97-119048/11.
DR N-PSDB: T51253.
PT New Alzheimer's disease related gene, AD4 - used to develop prods.
PT for detecting pre-disposition to or for diagnosis, prevention or
PT treatment of Alzheimer's disease
PS Example 4; Fig 2; 83pp; English.
CC A genetically isolated group of families with autosomal dominant
CC early-onset Alzheimer's Disease (AD) has been studied and initial
CC mapping analyses have predicted the AD locus resides on chromosome 1.
CC The present sequence corresponds to the protein encoded by the
CC wild-type coding sequence (i.e. the AD4 gene, also known as STM2).
CC The group of families has been designated the Volga German (VG)
CC kindreds. The entire gene has been amplified from VG individuals
CC and unaffected individuals (from VG and unrelated lineages).
CC Sequence analysis has shown that affected individuals have a
CC nucleotide change at codon 141 resulting in an amino acid alteration
CC from Asn to Ile. Portions of a mutant AD4, especially one in which
CC Asn at position 141 has been replaced by Ile, can be used in a
CC peptide vaccine. Detection of mutant AD4, for example using antibodies
CC specific for the protein or using nucleic acid probes specific for the

CC mutant gene, provides a means of diagnosing Alzheimer's disease.
SQ Sequence 448 AA;

Query Match 65.2%; Score 43; DB 24; Length 448;
Best Local Similarity 55.6%; Pred. No. 1.97e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 30 scqegrqgp 38
||:|:|:|
Qy 554 SCEPGHOSP 562

Search completed: Thu Jul 8 19:05:43 1999
Job time : 31 secs.

THIS PAGE BLANK (USPTO)

MPERCH_PP protein - protein database search, using Smith-Waterman algorithm

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Thu Jul 8 19:04:42 1999; MasPar time 5.08 Seconds
Tabular output not generated. 70.996 Million cell updates/sec

Title: >US-09-041-236-2
Description: (554-562) from US09041236.pep (28 of 45)
Perfect Score: 66
Sequence: 1 SCEPGHQSP 9

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 20.008; Variance 24.434; scale 0.819

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	49	74.2	35	2	JC6157 viral capsid protein	9.01e-01
2	49	74.2	769	1	JC1121 leukocyte adhesion pr	9.01e-01
3	47	71.2	577	2	B37057 integrin beta-6 chain	2.50e+00
4	46	69.7	788	2	A37057 integrin beta-6 chain	4.13e+00
5	45	68.2	386	2	E71001 probable serine amino	6.77e+00
6	45	68.2	445	2	A24701 tubulin beta-3 chain	6.77e+00
7	45	68.2	454	2	A38643 protein kinase (EC 2.	6.77e+00
8	45	68.2	484	2	S53641 protein kinase ckl1	6.77e+00
9	45	68.2	510	2	A42750 insulinoma-associated	6.77e+00
10	44	66.7	235	2	F64799 hypothetical protein	1.10e+01
11	44	66.7	295	1	SL0092 homeotic protein Hox	1.10e+01
12	44	66.7	295	2	JE0051 SCOL protein precurs	1.10e+01
13	44	66.7	301	2	S45880 SCOL protein homolog	1.10e+01
14	44	66.7	311	2	A64800 purine nucleosidase-r	1.10e+01
15	44	66.7	313	2	A64985 hypothetical 33.7 kD	1.10e+01
16	44	66.7	352	2	C70532 hypothetical protein	1.10e+01
17	44	66.7	378	2	S69883 hypothetical protein	1.10e+01
18	44	66.7	399	2	E33282 DNA-binding protein (1.10e+01
19	44	66.7	417	2	S45570 nuclear factor I-C4	1.10e+01
20	44	66.7	439	2	S09998 nuclear factor I-C2	1.10e+01
21	44	66.7	510	2	S45571 nuclear factor I-C1	1.10e+01
22	44	66.7	918	2	S20550 sensor-regulator prot	1.10e+01
23	44	66.7	1204	1	DNB229 DNA-binding protein -	1.10e+01

24	43	65.2	132	2	PH1289 Ig heavy chain precu	1.78e+01
25	43	65.2	306	2	A48993 protein-tyrosine-phos	1.78e+01
26	43	65.2	354	1	S10065 L-iditol 2-dehydrogen	1.78e+01
27	43	65.2	357	1	A54674 L-iditol 2-dehydrogen	1.78e+01
28	43	65.2	375	1	S54458 alcohol dehydrogenase	1.78e+01
29	43	65.2	442	2	I39174 seven trans-membrane	1.78e+01
30	43	65.2	448	2	I58098 E5-1 protein - human	1.78e+01
31	43	65.2	448	2	A56993 presenilin 2 - human	1.78e+01
32	43	65.2	657	2	S32739 filensin - chicken	1.78e+01
33	43	65.2	982	2	S16292 nitrate reductase (NA	1.78e+01
34	43	65.2	1378	2	I48751 protein-tyrosine kina	1.78e+01
35	43	65.2	1442	2	B48148 protein-tyrosine-phos	1.78e+01
36	43	65.2	2871	2	A55624 fibrillin-1 precursor	1.78e+01
37	43	65.2	3002	2	A47221 fibrillin 1 precursor	1.78e+01
38	42	63.6	236	1	XMBSTA transcription activat	2.86e+01
39	42	63.6	375	1	S65956 L-iditol 2-dehydrogen	2.86e+01
40	42	63.6	396	1	KXBOZ plasma protein 2 - bo	2.86e+01
41	42	63.6	399	1	S16132 L-iditol 2-dehydrogen	2.86e+01
42	42	63.6	435	2	A38153 paired box protein Po	2.86e+01
43	42	63.6	1690	1	CGHUIB collagen alpha 4(IV)	2.86e+01
44	42	63.6	1691	1	S22917 collagen alpha 5(IV)	2.86e+01
45	42	63.6	2907	2	A57278 fibrillin-2 precursor	2.86e+01

ALIGNMENTS

RESULT 1
ENTRY JC6157 #type complete
TITLE viral capsid protein 35 - Human herpesvirus 8, Kaposi
ORGANISM sarcoma-associated herpesvirus
#formal_name Human herpesvirus 8, Kaposi sarcoma-associated
herpesvirus
DATE 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change
05-Sep-1997
ACCESSIONS JC6157
REFERENCE JC6153
#authors Zhong, W.; Wang, H.; Herndier, B.; Ganem, D.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:6641-6646
#title Restricted expression of Kaposi sarcoma-associated
herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma.
#cross-references MUID:96270595
#accession JC6157
#molecule_type mRNA
#residues 1-35 #label ZHO
#cross-references GB:U75698
COMMENT This protein is a marker for the subpopulation involved in
productive infection or in abortive versions of the lytic cycle.
KEYWORDS membrane protein
SUMMARY #length 35 #molecular-weight 3831 #checksum 8586
Query Match 74.2%; Score 49; DB 2; Length 35;
Best Local Similarity 62.5%; Pred. No. 9.01e-01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 26 CQPGHSS 33
:||||:
QY 555 CEPGHQSP 562
#length 35 #molecular-weight 3831 #checksum 8586

RESULT 2
ENTRY JC1121 #type complete
TITLE leukocyte adhesion protein beta chain precursor - bovine
ALTERNATE_NAMES surface glycoprotein CD18
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
05-Sep-1997
ACCESSIONS JC1121
REFERENCE JC1121
#authors Shuster, D.E.; Bosworth, B.T.; Kehrl Jr., M.E.
#journal Gene (1992) 114:267-271
#title Sequence of the bovine CD18-encoding cDNA: Comparison with
the human and murine glycoproteins.
#cross-references MUID:92290287


```
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 307 EPGHESP 313
QY 556 EPGHQP 562
|||||
Query Match 68.2%; Score 45; DB 2; Length 454;
Best Local Similarity 71.4%; Pred. No. 6.77e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 6
ENTRY A24701 #type complete
TITLE tubulin beta-3 chain - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change
20-Mar-1998
ACCESSIONS A24701; A27554
REFERENCE A24701
#authors Sullivan, K.F.; Machlin, P.S.; Ratrie III, H.; Cleveland, D.W.
#journal J. Biol. Chem. (1986) 261:13317-13322
#title Sequence and expression of the chicken beta-3-tubulin gene. A vertebrate testis beta-tubulin isotype.
#cross-references MUID:87008546
#accession A24701
#molecule_type DNA
#residues 1-445 #label SUL
#cross-references GB:M14228; NID:g212829; PID:g212830
REFERENCE A27554
#authors Sullivan, K.F.; Cleveland, D.W.
#journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:4327-4331
#title Identification of conserved isotype-defining variable region sequences for four vertebrate beta-tubulin polypeptide classes.
#cross-references MUID:86233423
#accession A27554
#molecule_type DNA
#residues 1-27,'R','R',29-81,'GR',84-306,'D','308-444','F' #label SU2
#cross-references EMBL:M11443
#superfamily tubulin
CLASSIFICATION #length 445 #molecular-weight 49861 #checksum 4538
SUMMARY

Query Match 68.2%; Score 45; DB 2; Length 445;
Best Local Similarity 83.3%; Pred. No. 6.77e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 302 ACEPGH 307
QY 554 SCEPGH 559
|||||

RESULT 7
ENTRY A38643 #type complete
TITLE protein kinase (EC 2.7.1.37) cdc2/cdc28-like - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Feb-1992 #sequence_revision 21-Feb-1992 #text_change
10-Jul-1998
ACCESSIONS A38643
REFERENCE A38643
#authors Johnson, K.W.; Smith, K.A.
#journal J. Biol. Chem. (1991) 266:3402-3407
#title Molecular cloning of a novel human cdc2/CDC28-like protein kinase.
#cross-references MUID:91139618
#accession A38643
#status preliminary
#molecule_type mRNA
#residues 1-454 #label JOH
#cross-references GB:M59287
GENETICS
#gene GDB:CLK
#cross-references GDB:I26862
#map_position 7q31-7q31
CLASSIFICATION #superfamily protein kinase homology
KEYWORDS phosphotransferase
FEATURE

129-400 #domain protein kinase homology #label KIN
SUMMARY #length 454 #molecular-weight 53709 #checksum 2284

Query Match 68.2%; Score 45; DB 2; Length 454;
Best Local Similarity 71.4%; Pred. No. 6.77e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 83 GCEPGHR 89
QY 554 SCEPGHQ 560
|||||

RESULT 8
ENTRY S53641 #type complete
TITLE protein kinase clk1 (EC 2.7.1.-) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
10-Jul-1998
ACCESSIONS S53641
REFERENCE S53637
#authors Hanes, J.; von der Kammer, H.; Klaudiny, J.; Scheit, K.H.
#journal J. Mol. Biol. (1994) 244:665-672
#title Characterization by cDNA cloning of two new human protein kinases. Evidence by sequence comparison of a new family of mammalian protein kinases.
#cross-references MUID:95082033
#accession S53641
#status nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-484 #label HAN
#cross-references GB:L29219; NID:9632963; PID:9632964
CLASSIFICATION #superfamily protein kinase homology
KEYWORDS alternative splicing; phosphotransferase
FEATURE 159-430
SUMMARY #domain protein kinase homology #label KIN
#length 484 #molecular-weight 57205 #checksum 3358

Query Match 68.2%; Score 45; DB 2; Length 484;
Best Local Similarity 71.4%; Pred. No. 6.77e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 83 GCEPGHR 89
QY 554 SCEPGHQ 560
|||||

RESULT 9
ENTRY A42750 #type complete
TITLE insulinoma-associated protein IA-1 - human
ALTERNATE_NAMES probable neuroendocrine cell transformation protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
17-Mar-1999
ACCESSIONS A42750; A54088
REFERENCE A42750
#authors Goto, Y.; De Silva, M.G.; Toscani, A.; Prabhakar, B.S.; Notkins, A.L.; Lan, M.S.
#journal J. Biol. Chem. (1992) 267:15252-15257
#title A novel human insulinoma-associated cDNA, IA-1, encodes a protein with "zinc-finger" DNA-binding motifs.
#cross-references MUID:92340582
#accession A42750
#status preliminary
#molecule_type mRNA
#residues 1-510 #label GOT
#cross-references GB:M93119; NID:9184510; PID:g184511
#experimental_source insulinoma cells
#note sequence extracted from NCBI backbone (NCBIN:109495, NCBIP:109496)
REFERENCE A54088
#authors Lan, M.S.; Li, Q.; Lu, J.; Modi, W.S.; Notkins, A.L.
#journal J. Biol. Chem. (1994) 269:14170-14174
#title Genomic organization, 5'-upstream sequence, and chromosomal
```

```
localization of an insulinoma-associated intronless gene,
IA-1.
#cross-references MUID:94245740
#accession A54088
##status preliminary
##molecule_type DNA
##residues 1-11 #label LAN
#cross-references GB:007172; NID:g488074; PID:g531163
GENETICS
#gene GDB:INSM1
#map_position 20p11.2-20p11.2
#status absent
#introns
#keywords DNA binding; transcription regulation; zinc finger
SUMMARY
#length 510 #molecular-weight 52923 #checksum 5956

Query Match 68.2%; Score 45; DB 2; Length 510;
Best Local Similarity 55.6%; Pred. NO. 6.77e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 71 ACAPGQPP 79
:| | | |
QY 554 SCEPGHQP 562

RESULT 10
ENTRY F64799 #type complete
TITLE hypothetical protein b0648 - Escherichia coli
ALTERNATE_NAMES #formal_name Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
17-Jul-1998
ACCESSIONS F64799
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession F64799 nucleic acid sequence not shown; translation not shown
##status
##molecule_type DNA
#residues 1-235 #label BLAT
#cross-references GB:AE000169; GB:U00096; NID:g1786862; PID:g1786868;
#experimental_source strain K-12, substrain MG1655
SUMMARY #length 235 #molecular-weight 27022 #checksum 349

Query Match 66.7%; Score 44; DB 2; Length 235;
Best Local Similarity 55.6%; Pred. NO. 1.10e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 22 GCKPAHDS 30
:| | | |
QY 554 SCEPGHQP 562

RESULT 11
ENTRY S10092 #type complete
TITLE homeotic protein Hox B4 - chicken
ALTERNATE_NAMES homeotic protein Chox-2.6; homeotic protein Chox-Z
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 02-Dec-1993 #sequence_revision 03-Aug-1995 #text_change
05-Sep-1997
ACCESSIONS S10092
REFERENCE S10092
#authors Sasaki, H.; Kuroiwa, A.
#journal Nucleic Acids Res. (1990) 18:184
#title The nucleotide sequence of the cDNA encoding a chicken
deformed family homeobox gene, Chox-Z.
#cross-references MUID:90174917
```

```
#accession S10092
#status preliminary
##molecule_type mRNA
##residues 1-245 #label SAS
#cross-references EMBL:X17612; NID:g63504; PID:g63505
FUNCTION #description control of embryonic development by tissue- and
stage-specific regulation of transcription
CLASSIFICATION #superfamily homeotic protein Hox D4; homeobox homology
KEYWORDS DNA binding; embryo; homeobox; nucleus; transcription
regulation
FEATURE
151-207 #domain homeobox homology #label HOX
SUMMARY #length 245 #molecular-weight 27782 #checksum 7772

Query Match 66.7%; Score 44; DB 1; Length 245;
Best Local Similarity 66.7%; Pred. NO. 1.10e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 98 PCFPGTSP 106
:| | | |
QY 554 SCEPGHQP 562

RESULT 12
ENTRY JE0051 #type complete
TITLE SCO1 protein precursor - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YBR037c; protein YBR0406
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change
06-Feb-1998
ACCESSIONS JE0051; S45895
REFERENCE JE0051
#authors Schulze, M.; Roedel, G.
#journal Mol. Gen. Genet. (1989) 216:37-43
#title Accumulation of the cytochrome c oxidase subunits I and II in
yeast requires a mitochondrial membrane-associated protein,
encoded by the nuclear SCO1 gene.
#cross-references MUID:89281488
#accession JE0051
##molecule_type DNA
#residues 1-295 #label SCH
#cross-references EMBL:X17441; NID:g4427; PID:g666104
REFERENCE S45893
#authors Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.;
Urrestarazu, A.; Vissers, S.
#submission submitted to the Protein Sequence Database, August 1994
#accession S45895
##molecule_type DNA
#residues 1-295 #label AND
#cross-references EMBL:Z35906; NID:g536255; PID:g536256; MIPS:YBR037c
GENETICS
#gene SGD:SCO1
#cross-references SGD:S0000241; MIPS:YBR037c
#map_position 2R
#genome nuclear
KEYWORDS mitochondrion; transmembrane protein
FEATURE
75-92 #domain transmembrane #status predicted #label TMM
SUMMARY #length 295 #molecular-weight 33166 #checksum 7834

Query Match 66.7%; Score 44; DB 2; Length 295;
Best Local Similarity 44.4%; Pred. NO. 1.10e+01;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

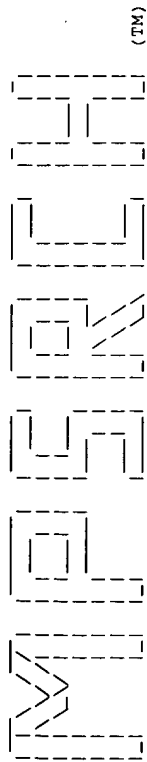
Db 180 TCDPARDSP 188
:| | | |
QY 554 SCEPGHQP 562

RESULT 13
ENTRY S45880 #type complete
TITLE SCO1 protein homolog SCO2 - yeast (Saccharomyces cerevisiae)
```

ALTERNATE_NAMES hypothetical protein YBR024w: hypothetical protein YBR0308
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change
 06-Feb-1998
ACCESSIONS S45880; S46555
REFERENCE S45875
#authors Grivell, L.A.; de Haan, M.; Maat, M.J.; Smits, P.H.M.
#submission Submitted to the Protein Sequence Database, August 1994
#accession S45880
 #molecule_type DNA
 #residues 1-301 #label GRI
 #cross-references EMBL:Z35893; NID:g536231; MIPS:YBR024w
 #experimental_source strain S288C
REFERENCE S46551
#authors Yeats, P.H.M.; de Haan, M.; Maat, C.; Grivell, L.A.
#journal Yeast (1994) 10(Suppl.A):S75-S80
#title The complete sequence of a 33 kb fragment on the right arm of
 chromosome II from Saccharomyces cerevisiae reveals 16 open
 reading frames, including ten new open reading frames, five
 previously identified genes and a homologue of the SCO1
 gene.
#accession S46555
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-301 #label SMT
#cross-references EMBL:X76078; NID:g498748; PID:g498753
#experimental_source strain S288C
#note The nucleotide sequence was submitted to the EMBL Data
 Library, November 1993
GENETICS
#gene SGD:SCO2
#cross-references SGD:S0000228; MIPS:YBR024w
#map_position 2R
KEYWORDS mitochondrion; transmembrane protein
SUMMARY #length 301 #molecular-weight 34888 #checksum 1025
 66.7%; Score 44; DB 2; Length 301;
Query Match Best Local Similarity 44.4%; Pred. No. 1.10e+01;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 186 SCDPARDTP 194
 11111111
Qy 554 SCEPGHQSP 562
 10 CDPGHDDA 17
 11111111
RESULT 14
ENTRY purine nucleosidase-related protein ybeK - Escherichia coli
TITLE #formal_name Escherichia coli
ORGANISM 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
DATE 18-Sep-1998
ACCESSIONS A64800
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
 Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
 Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
 Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
 Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:9742617
#accession A64800
#status preliminary; nucleic acid sequence not shown;
 translation not shown
#molecule_type DNA
#residues 1-313 #label BLAT
#cross-references GB:AE000305; GB:U00096; NID:gl788479; PID:gl788486;
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene ybeK
CLASSIFICATION #superfamily yaaF protein
SUMMARY #length 313 #molecular-weight 33748 #checksum 5179
 66.7%; Score 44; DB 2; Length 313;
Query Match Best Local Similarity 50.0%; Pred. No. 1.10e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 10 CDPGHDDA 17
 11111111
Qy 555 CEPGHQSP 562
 Search completed: Thu Jul 8 19:04:53 1999
 Job time : 11 secs.

ALTERNATE_NAMES hypothetical protein YBR024w: hypothetical protein YBR0308
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change
 06-Feb-1998
ACCESSIONS S45880; S46555
REFERENCE S45875
#authors Grivell, L.A.; de Haan, M.; Maat, M.J.; Smits, P.H.M.
#submission Submitted to the Protein Sequence Database, August 1994
#accession S45880
 #molecule_type DNA
 #residues 1-301 #label GRI
 #cross-references EMBL:Z35893; NID:g536231; MIPS:YBR024w
 #experimental_source strain S288C
REFERENCE S46551
#authors Yeats, P.H.M.; de Haan, M.; Maat, C.; Grivell, L.A.
#journal Yeast (1994) 10(Suppl.A):S75-S80
#title The complete sequence of a 33 kb fragment on the right arm of
 chromosome II from Saccharomyces cerevisiae reveals 16 open
 reading frames, including ten new open reading frames, five
 previously identified genes and a homologue of the SCO1
 gene.
#accession S46555
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-301 #label SMT
#cross-references EMBL:X76078; NID:g498748; PID:g498753
#experimental_source strain S288C
#note The nucleotide sequence was submitted to the EMBL Data
 Library, November 1993
GENETICS
#gene SGD:SCO2
#cross-references SGD:S0000228; MIPS:YBR024w
#map_position 2R
KEYWORDS mitochondrion; transmembrane protein
SUMMARY #length 301 #molecular-weight 34888 #checksum 1025
 66.7%; Score 44; DB 2; Length 301;
Query Match Best Local Similarity 44.4%; Pred. No. 1.10e+01;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 186 SCDPARDTP 194
 11111111
Qy 554 SCEPGHQSP 562
 10 CDPGHDDA 17
 11111111
RESULT 14
ENTRY purine nucleosidase-related protein ybeK - Escherichia coli
TITLE #formal_name Escherichia coli
ORGANISM 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
DATE 18-Sep-1998
ACCESSIONS A64800
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
 Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
 Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
 Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
 Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:9742617
#accession A64800
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-311 #label BLAT
#cross-references GB:AE000169; GB:U00096; NID:gl786862; PID:gl786871;
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene ybeK
CLASSIFICATION #superfamily yaaF protein
KEYWORDS transmembrane protein
FEATURE

THIS PAGE BLANK (USPTO)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:03:36 1999; MasPar time 3.36 Seconds
Tabular output not generated. 75.687 Million cell updates/sec

Title: >US-09-041-236-2
Description: (554-562) from US09041236.pep (28 of 45)
Perfect Score: 66
Sequence: 1 SCEPGHOSP 9

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 20.649; Variance 22.790; scale 0.906

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	80.3	404	1	GAT5_MOUSE TRANSCRIPTION FACTOR G	3.69e-02
2	49	74.2	769	1	ITB2_BOVIN CELL SURFACE ADHESION	3.60e-01
3	47	71.2	577	1	ITB6_CAVPO INTEGRIN BETA-6 SUBUNIT	1.08e+00
4	46	69.7	788	1	ITB6_HUMAN INTEGRIN BETA-6 SUBUNIT	1.85e+00
5	45	68.2	95	1	Y145_NPVOP HYPOTHETICAL 11.0 KD P	3.15e+00
6	45	68.2	445	1	TBB3_CHICK TUBULIN BETA-3 CHAIN (3.15e+00
7	45	68.2	484	1	CLK1_HUMAN PROTEIN KINASE CLK1 (E	3.15e+00
8	45	68.2	510	1	IAT1_HUMAN ZINC FINGER PROTEIN IA	3.15e+00
9	45	68.2	587	1	NDC2_RAT INTESTINAL SODIUM/DICA	3.15e+00
10	44	66.7	235	1	YBEU_ECOLI HYPOTHETICAL 27.0 KD P	5.32e+00
11	44	66.7	245	1	HXB4_CHICK HOMEBOX PROTEIN HOX-B	5.32e+00
12	44	66.7	295	1	SC01_YEAST SC01 PROTEIN PRECURSOR	5.32e+00
13	44	66.7	301	1	SC02_YEAST SC02 PROTEIN PRECURSOR	5.32e+00
14	44	66.7	311	1	YBEK_ECOLI HYPOTHETICAL 33.8 KD P	5.32e+00
15	44	66.7	313	1	YDIK_ECOLI HYPOTHETICAL 33.7 KD P	5.32e+00
16	44	66.7	378	1	YD40_YEAST HYPOTHETICAL 42.3 KD P	5.32e+00
17	44	66.7	388	1	VE2_HPV29 REGULATORY PROTEIN E2.	5.32e+00
18	44	66.7	439	1	NFIC_CHICK NUCLEAR FACTOR 1 C2 (N	5.32e+00
19	44	66.7	537	1	Z029_XENIA OOCYTE ZINC FINGER PRO	5.32e+00
20	44	66.7	918	1	BARA_ECOLI SENSOR PROTEIN BARA (E	5.32e+00
21	44	66.7	1204	1	DNBI_VZVD MAJOR DNA-BINDING PROT	5.32e+00
22	43	65.2	330	1	YD62_SCHPO HYPOTHETICAL 36.1 KD P	8.91e+00
23	43	65.2	354	1	DHSO_SHEEP SORBITOL DEHYDROGENASE	8.91e+00

24	43	65.2	356	1	DHSO_HUMAN SORBITOL DEHYDROGENASE	8.91e+00
25	43	65.2	375	1	ADH3_YEAST ALCOHOL DEHYDROGENASE	8.91e+00
26	43	65.2	448	1	PSN2_HUMAN PRESENILIN 2 (PS-2) (S	8.91e+00
27	43	65.2	657	1	FILS_CHICK FLENSILIN (LENS FIBER C	8.91e+00
28	43	65.2	982	1	NIA_NEUCR NITRATE REDUCTASE (NAD	8.91e+00
29	43	65.2	1442	1	PTPG_MOUSE PROTEIN-TYROSINE PHOSP	8.91e+00
30	43	65.2	2871	1	FBN1_HUMAN FIBRILLIN 1 PRECURSOR.	8.91e+00
31	43	65.2	2871	1	FBN1_MOUSE FIBRILLIN 1 PRECURSOR.	8.91e+00
32	42	63.6	236	1	TENA_BACSU TRANSCRIPTIONAL ACTIVA	1.48e+01
33	42	63.6	375	1	DHSO_MOUSE SORBITOL DEHYDROGENASE	1.48e+01
34	42	63.6	396	1	PRIZ_BOVIN VITAMIN K-DEPENDENT PR	1.48e+01
35	42	63.6	399	1	DHSO_RAT SORBITOL DEHYDROGENASE	1.48e+01
36	42	63.6	425	1	POXN_DROME PAIRED BOX POX-NEURO P	1.48e+01
37	42	63.6	451	1	TBAL_HUMAN TUBULIN ALPHA-1 CHAIN,	1.48e+01
38	42	63.6	1885	1	CAS4_HUMAN COLLAGEN ALPHA 5(IV) C	1.48e+01
39	42	63.6	1690	1	CA44_HUMAN COLLAGEN ALPHA 4(IV) C	1.48e+01
40	42	63.6	1964	1	NTC4_MOUSE NEUROGENIC LOCUS NOTCH	1.48e+01
41	42	63.6	2871	1	FBN1_BOVIN FIBRILLIN 1 PRECURSOR	1.48e+01
42	42	63.6	2907	1	FBN2_MOUSE FIBRILLIN 2 PRECURSOR.	1.48e+01
43	42	63.6	2911	1	FBN2_HUMAN FIBRILLIN 2 PRECURSOR.	1.48e+01
44	41	62.1	516	1	ZG53_XENLA GASTRULA ZINC FINGER P	2.43e+01
45	41	62.1	1313	1	ACE_RAT ANGIOTENSIN-CONVERTING	2.43e+01

ALIGNMENTS

RESULT 1
ID GAT5_MOUSE STANDARD; PRT: 404 AA.
AC P97489;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION FACTOR GATA-5 (GATA BINDING FACTOR-5).
GN GAT5.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA MORRISSEY E., IP H.S., LU M.M., PARMACEK M.S.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: TO OTHER GATA-TYPE TRANSCRIPTION FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U84725; G1841756; -
CC MGD: MGI:109497; GATA5.
CC PROSITE: PS00344; GATA_2N_FINGER; 2.
CC PFAM: PF00320; GATA; 2.
CC HSSP: P17678; 1GAU.
CC TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; ZINC-FINGER;
CC NUCLEAR PROTEIN.
FT ZN_FING 196 220 GATA-TYPE.
FT ZN_FING 250 274 GATA-TYPE.
SQ SEQUENCE 404 AA: 42145 MW; 8B7421EE CRC32;

Query Match 80.3%; Score 53; DB 1; Length 404;
Best Local Similarity 77.8%; Pred. No. 3.69e-02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 48 SCEPGSQAP 56
QY 554 SCEPGHOSP 562

RESULT 2
ID ITB2_BOVIN STANDARD; PRT; 769 AA.
AC P32592;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE CELL SURFACE ADHESION GLYCOPROTEINS LFA-1/CR3/PL150.95 BETA-SUBUNIT
DE PRECURSOR (INTEGRIN BETA-2) (CD18 ANTIGEN) (COMPLEMENT RECEPTOR C3
DE BETA-SUBUNIT).
GN ITGB2 OR CD18.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92290287.
RA SHUSTER D.E., BOSWORTH B.T., KEHLRI M.E. JR.;
RT "Sequence of the bovine CD18-encoding cDNA: comparison with the human
RT and murine glycoproteins.";
RL GENE 114:267-271(1992).
RN [2]
RP VARIANT LAD GLY-128.
RX MEDLINE: 93028437.
RA SHUSTER D.E., KEHLRI M.E. JR., ACKERMANN M.R., GILBERT R.O.;
RT "Identification and prevalence of a genetic defect that causes
RT leukocyte adhesion deficiency in Holstein cattle.";
RL PROC. NATL. ACAD. SCI. U.S.A. 89:9225-9229(1992).
CC -!- FUNCTION: ASSOCIATES WITH ALPHA-L (LFA-1) TO INTERACT WITH ICAM-1.
CC AND WITH ALPHA-M (MAC-1) OR ALPHA-X TO FORM THE RECEPTOR FOR THE
CC IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT.
CC -!- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-2 ASSOCIATES
CC WITH THE ALPHA-D, -L, -M AND -X CHAINS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- PTM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE
CC BONDS.
CC -!- DISEASE: DEFECTS IN CD18 ARE THE CAUSE OF LEUKOCYTE ADHESION
CC DEFICIENCY (LAD). THE MUTATION CAUSING LAD (GLY-128) IS PREVALENT
CC AMONG HOLSTEIN CATTLE THROUGHOUT THE WORLD, PLACING THIS DISORDER
CC AMONG THE MOST COMMON GENETIC DISEASES KNOWN IN ANIMAL
CC AGRICULTURE. ALL CATTLE WITH THE MUTANT ALLELE ARE RELATED TO ONE
CC BULL, WHO THROUGH THE USE OF ARTIFICIAL INSEMINATION Sired MANY
CC CALVES IN THE 1950S AND 1960S.
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M81233; G162819; -.
CC PIR: JC1121; JC1121.
CC PROSITE: PS00243; INTEGRIN_BETA; 3.
CC PROSITE: PS00022; EGF_1; UNKNOWN_2.
CC PROSITE: PS01186; EGF_2; UNKNOWN_2.
CC PFAM: PF00362; integrin_B; 1.
CC INTEGRIN; CELL ADHESION; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT;
CC EXTRACELLULAR MATRIX; CYTOSKELETON; SIGNAL; DISEASE MUTATION.
CC SIGNAL 1 22
CC CHAIN 23 769
CC DOMAIN 23 700
CC TRANSMEM 701 723
CC DOMAIN 724 769
CC DOMAIN 449 617
CC REPEAT 497 540
CC REPEAT 541 581
CC REPEAT 582 617
CC SITE 397 399
CC CARBOHYD 50 50
CC CARBOHYD 116 116

FT CARBOHYD 254 254 POTENTIAL.
FT CARBOHYD 501 501 POTENTIAL.
FT CARBOHYD 642 642 POTENTIAL.
FT VARIANT 128 128 D -> G (IN LAD).
SQ SEQUENCE 769 AA; 84400 MW; 46F481EE CRC32;
Query Match 74.2%; Score 49; DB 1; Length 769;
Best Local Similarity 62.5%; Pred. No. 3.60e-01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 603 CDPGYQPP 610
QY 555 CEPGHQSP 562
RESULT 3
ID ITB6_CAVPO STANDARD; PRT; 577 AA.
AC P18563;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE INTEGRIN BETA-6 SUBUNIT (FRAGMENT).
GN ITGB6.
OS CAVIA PORCELLUS (GUINEA PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; HYSTRICOGNATHI; CAVIIDAE; CAVIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HARTLEY;
RX MEDLINE: 90307659.
RA SHEPPARD D., ROZZO C., STARR L., QUARANTA V., ERLE D.J., PYTELA R.;
RT "Complete amino acid sequence of a novel integrin beta subunit (beta
RT 6) identified in epithelial cells using the polymerase chain
RT reaction.";
RL J. BIOL. CHEM. 265:11502-11507(1990).
CC -!- FUNCTION: INTEGRINS ARE A LARGE FAMILY OF CELL SURFACE
CC GLYCOPROTEINS THAT MEDIANTE CELL TO CELL & CELL TO MATRIX ADHESION.
CC -!- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-6 ASSOCIATES
CC WITH ALPHA-V.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- PTM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE
CC BONDS.
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M35197; G553845; -.
CC PIR: B37057; B37057.
CC PROSITE: PS00243; INTEGRIN_BETA; 2.
CC PROSITE: PS00022; EGF_1; UNKNOWN_2.
CC PROSITE: PS01186; EGF_2; UNKNOWN_1.
CC PFAM: PF00362; integrin_B; 1.
CC HSP; P04355; 2MRT.
CC INTEGRIN; CELL ADHESION; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT;
CC EXTRACELLULAR MATRIX; CYTOSKELETON.
CC NON_TER 1 1
CC DOMAIN <1 566 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 567 >577 POTENTIAL.
CC DOMAIN 315 478 CYSTEINE-RICH REPEATS.
CC REPEAT 315 360 I.
CC REPEAT 361 402 II.
CC REPEAT 403 441 III.
CC REPEAT 442 478 IV.
CC CARBOHYD 119 119 POTENTIAL.
CC CARBOHYD 246 246 POTENTIAL.
CC CARBOHYD 255 255 POTENTIAL.

FT CARBOHYD 277 277 POTENTIAL.
FT CARBOHYD 322 322 POTENTIAL.
FT CARBOHYD 330 330 POTENTIAL.
FT CARBOHYD 400 400 POTENTIAL.
FT CARBOHYD 434 434 POTENTIAL.
FT CARBOHYD 555 555 POTENTIAL.
FT NON_TER 577 577
SQ SEQUENCE 577 AA; 62298 MW; AE102507 CRC32;

Query Match 71.2%; Score 47; DB 1; Length 577;
Best Local Similarity 55.6%; Pred. No. 1.08e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 339 ACNPGHMG 347
QY 554 SCEPGHQP 562
:|||||:

RESULT 4
ID ITB6_HUMAN STANDARD; PRT; 788 AA.
AC P18564; Q16500;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DE INTEGRIN BETA-6 SUBUNIT (UPDATE)
DE INTEGRIN BETA-6 SUBUNIT PRECURSOR.
GN ITGB6
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PANCREAS;
RX MEDLINE; 90307659.
RA SHEPPARD D., ROZZO C., STARR L., QUARANTA V., ERLE D.J., PYTELA R.;
RT "Complete amino acid sequence of a novel integrin beta subunit (beta
6) identified in epithelial cells using the polymerase chain
reaction.";
RT J. BIOL. CHEM. 265:11502-11507(1990).
RN [2]
RP SEQUENCE OF 116-197 FROM N.A.
RX MEDLINE; 93002753.
RA JIANG W.M., JENKINS D., YUAN Q., LEUNG E., CHOO K.H., WATSON J.D.,
RA KRISANSSEN G.W.;
RT "The gene organization of the human beta 7 subunit, the common beta
subunit of the leukocyte integrins HML-1 and LPAM-1.";
RL INT. IMMUNOL. 4:1031-1040(1992).
CC -!- FUNCTION: INTEGRINS ARE A LARGE FAMILY OF CELL SURFACE
GLYCOPROTEINS THAT MEDIATE CELL TO CELL & CELL TO MATRIX ADHESION.
CC -!- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-6 ASSOCIATES
CC WITH ALPHA-V.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- PTM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE
BONDS.
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M35198; G186507; -
CC EMBL; A26609; E199078; -
CC EMBL; S49380; G257589; -
CC PIR; A37057; A37057.
CC MIM; 147536; -
CC PROSITE; PS00243; INTEGRIN_BETA; 3.
CC PROSITE; PS00022; EGF_1; UNKNOWN.2.
CC PROSITE; PS01186; EGF_2; UNKNOWN.1.
CC PFAM; PF00362; integrin_B; 1.
CC HSSP; P04355; 2MRT.

KW INTEGRIN: CELL ADHESION; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT;
KW EXTRACELLULAR MATRIX; CYTOSKELETON; SIGNAL.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 788 INTEGRIN BETA-6.
FT DOMAIN 19 707 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 708 730 POTENTIAL.
FT DOMAIN 731 788 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 456 619 CYSTEINE-RICH REPEATS.
FT REPEAT 456 501 I.
FT REPEAT 502 543 II.
FT REPEAT 544 582 III.
FT REPEAT 583 619 IV.
FT CARBOHYD 48 48 POTENTIAL.
FT CARBOHYD 97 97 POTENTIAL.
FT CARBOHYD 260 260 POTENTIAL.
FT CARBOHYD 387 387 POTENTIAL.
FT CARBOHYD 396 396 POTENTIAL.
FT CARBOHYD 463 463 POTENTIAL.
FT CARBOHYD 471 471 POTENTIAL.
FT CARBOHYD 541 541 POTENTIAL.
FT CARBOHYD 575 575 POTENTIAL.
FT CONFLICT 158 158 G -> R (IN REF. 2).
SQ SEQUENCE 788 AA; 85975 MW; 81429BE0 CRC32;

Query Match 69.7%; Score 46; DB 1; Length 788;
Best Local Similarity 55.6%; Pred. No. 1.85e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 480 ACHPGHMG 488
QY 554 SCEPGHQP 562
:|||||:

RESULT 5
ID Y145_NPVOP STANDARD; PRT; 95 AA.
AC O10373;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 11.0 KD PROTEIN (ORF142).
OS ORGYIA PSEUDOTSUGATA MULTICAPSID POLYHEDROSIS VIRUS (OPMNPV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
OC NUCLEOPOLYHEDROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 9271300.
RA AHRENS C.H., RUSSELL R.R., FUNK C.J., EVANS J., HARWOOD S.,
RA ROHRMANN G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
polyhedrosis virus genome.";
RL VIROLOGY 229:381-399(1997).
CC -!- SIMILARITY: TO CORRESPONDING ORF IN ACNPNV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U75930; G1911388; -
CC HYPOTHETICAL PROTEIN.
KW SEQUENCE 95 AA; 10998 MW; 7137FC47 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 95;
Best Local Similarity 66.7%; Pred. No. 3.15e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 60 CDPGHE 65
QY 555 CEPGHQ 560
:|||||:

```

J. MOL. BIOL. 244:665-672(1994).
-!- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS
OF THE SPLICEOSOMAL COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF
REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA
SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES (BY
SIMILARITY).
-!- SUBCELLULAR LOCATION: NUCLEAR.
-!- PTM: AUTOPHOSPHORYLATES ON ALL THREE TYPE OF RESIDUES (BY
SIMILARITY).
-!- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
SPLICING, THE SHORTEST FORM LACK THE KINASE DOMAIN.
-!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
PROTEIN KINASES. BELONGS TO THE LAMMER SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
-----
EMBL; L29219; G632964; -
DR EMBL; L29222; G632962; -
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; Q16539; lwfc.
DR TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
KW TYROSINE-PROTEIN KINASE; PHOSPHORYLATION; NUCLEAR PROTEIN;
KW ALTERNATIVE SPLICING.
FT DOMAIN 161 477 PROTEIN KINASE.
FT NP_BIND 167 175 ATP (BY SIMILARITY).
FT BINDING 191 191 ATP (BY SIMILARITY).
FT ACT_SITE 288 288 BY SIMILARITY.
FT VARSPLIC 131 136 KSHRRK -> MKLLIL (IN SHORT FORM).
FT VARSPLIC 137 484 MISSING (IN SHORT FORM).
SQ SEQUENCE 484 AA; 57205 MW; 99B67338 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 484;
Best Local Similarity 71.4%; Pred. No. 3.15e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 83 GCEPGHR 89
QY 554 SCEPGHQ 560
:||||:

RESULT 8
ID IAL_HUMAN STANDARD; PRT; 510 AA.
AC Q01101.
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ZINC FINGER PROTEIN IA-1 (INSULINOMA-ASSOCIATED PROTEIN 1).
GN INSM1 OR IAL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92340582.
RA GOTO Y., DE SILVA M.G., TOSCANI A., PRABHAKAR B.S., NOTKINS A.L.,
RA LAN M.S.;
RT "A novel human insulinoma-associated cDNA, IA-1, encodes a protein
RT with 'zinc-finger' DNA-binding motifs.";
RL J. BIOL. CHEM. 267:15252-15257(1992).
RN [2]
RP SEQUENCE OF 1-11 FROM N.A.
RX MEDLINE; 94245740.
RA LAN M.S., LI Q., LU J., MODI W.S., NOTKINS A.L.;
RT "Genomic organization, 5'-upstream sequence, and chromosomal
```


Qy 554 SCEPGHOSP 562

```
RESULT 11
ID HXB4_CHICK STANDARD; PRT; 245 AA.
AC P14840;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE HOMEOBOX PROTEIN HOX-B4 (CHOX-Z).
GN HOXB4 OR CHOX-Z.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90174917.
RA SASAKI H., KUROIWA A.;
RT "The nucleotide sequence of the cDNA encoding a chicken Deformed
family homeobox gene, Chox-Z."
RL NUCLEIC ACIDS RES. 18:184-184(1990).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEOBOX PROTEINS.
CC BELONGS TO THE "DEFORMED" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17612; G63505; -.
CC PIR; S10092; S10092.
CC DR PROSITE; PS00027; HOMEOBOX_1; 1.
CC DR PROSITE; PS00032; ANTENNAPEDIA; 1.
CC DR PROSITE; PS50071; HOMEOBOX_2; 1.
CC DR PFAM; PF00046; homeobox; 1.
CC DR HSP; P02833; ISAN.
CC DR TRANSFAC; T01726; -.
CC KW HOMEOBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN;
CC TRANSCRIPTION REGULATION.
FT DOMAIN 129 134 ANTP-TYPE HEXAPEPTIDE.
FT DNA_BIND 150 209 HOMEOBOX.
SQ SEQUENCE 245 AA; 27782 MW; CF5D2064 CRC32;

Query Match 66.7%; Score 44; DB 1; Length 245;
Best Local Similarity 66.7%; Pred. No. 5.32e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

Db 98 PCEPGTSP 106

Qy 554 SCEPGHOSP 562

```
RESULT 12
ID SCOL_YEAST STANDARD; PRT; 295 AA.
AC P2833;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE SCOL PROTEIN PRECURSOR.
GN SCOL OR YBR037C OR YBR0406.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89281488.
```

```
RA SCHULZE M., ROEDEL G.;
RT "Accumulation of the cytochrome c oxidase subunits I and II in yeast
requires a mitochondrial membrane-associated protein, encoded by the
nuclear SCOL gene."
RL MOL. GEN. GENET. 216:37-43(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA ANDRE B., CZIEPLUCH C., HEIN C., JAUNIAUX J.C., URRESTARAZU A.,
RA VISSERS S.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE; 92049241.
RA BUCHWALD P., KRUMBECK G., ROEDEL G.;
RT "Immunological identification of yeast SCOL protein as a component of
the inner mitochondrial membrane."
RL MOL. GEN. GENET. 229:413-420(1991).
CC -!- FUNCTION: REQUIRED FOR THE ACCUMULATION OF SUBUNITS 1 AND 2 OF
CC CYTOCHROME C OXIDASE COMPLEX. POSSIBLY INVOLVED IN PROMOTING THE
CC ASSEMBLY OF THE SUBUNITS.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC -!- SIMILARITY: STRONG, TO SCO2.
CC -!- SIMILARITY: SOME, TO B.SUBTILIS YPMQ AND TO A.P.STUTZERI
CC HYPOTHETICAL PROTEIN IN APT 3'REGION (AC P47206).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17441; G666104; -.
CC DR EMBL; X17441; G666104; -.
CC DR PIR; JE0051; JE0051.
CC DR SGD; L0001815; SCOL.
CC KW MITOCHONDRION; TRANSMEMBRANE; TRANSIT PEPTIDE.
FT TRANSIT 1 295 MITOCHONDRION (POTENTIAL).
FT CHAIN 76 92 SCOL PROTEIN.
FT TRANSMEM 76 92 POTENTIAL.
SQ SEQUENCE 295 AA; 33166 MW; 571B9610 CRC32;

Query Match 65.7%; Score 44; DB 1; Length 295;
Best Local Similarity 44.4%; Pred. No. 5.32e+00;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
```

Db 180 TCDPARDSP 188

Qy 554 SCEPGHOSP 562

```
RESULT 13
ID SCOL_YEAST STANDARD; PRT; 301 AA.
AC P38072;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE SCOL PROTEIN PRECURSOR.
GN SCOL OR YBR024W OR YBR0308.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA SMITS P.H.M., DE HANN M., MAAT C., GRIVELL L.A.;
RT "The complete sequence of a 33 kb fragment on the right arm of
chromosome II from Saccharomyces cerevisiae reveals 16 open reading
frames, including ten new open reading frames, five previously
identified genes and a homologue of the SCOL gene."
RT
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE000169; G1786871; -;
 DR EMBL: U82598; G1778569; -;
 DR EMBL: U10981; -; NOT_ANNOTATED_CDS.
 DR ECOGENE: EG12701; YBEK.
 DR PROSITE: PS01247; IUNH; 1.
 DR PFAM: PF01156; IU_nuc_hydro; 1.
 DR HSP: Q27546; IMAS.
 KW HYPOTHETICAL PROTEIN; HYDROLASE; GLYCOSIDASE.
 FT ACT_SITE 240 240 BY SIMILARITY.
 SQ SEQUENCE 311 AA; 33823 MW; E4777FC4 CRC32;

Query Match 66.7%; Score 44; DB 1; Length 311;
 Best Local Similarity 50.0%; Pred. No. 5.32e+00;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 9 CDPGHDDA 16
 |::|::|:
 QY 555 CEFQHOSP 562

RESULT 15
 ID YEIK_ECOLI STANDARD; PRT; 313 AA.
 AC P33022;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 33.7 KD PROTEIN IN NFO-PRUA INTERGENIC REGION.
 GN YEIK.
 OS ESCHERICHIA COLI.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC ESCHERICHIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / BHB2600;
 RA RICHTERICH P., LAKEY N., GRYAN G., JAEHN L., MINTZ L., ROBISON K.,
 RA CHURCH G.M.;
 RL SUBMITTED (OCT-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL SCIENCE 277:1453-1474(1997).
 CC -!- SIMILARITY: BELONGS TO THE IUNH FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U00007; G405882; -;
 DR EMBL: AE000305; G1788486; -;
 DR ECOGENE: EG12030; YEIK.
 DR PROSITE: PS01247; IUNH; 1.
 DR PFAM: PF01156; IU_nuc_hydro; 1.
 DR HSP: Q27546; IMAS.
 KW HYPOTHETICAL PROTEIN; HYDROLASE; GLYCOSIDASE.
 FT ACT_SITE 239 239 BY SIMILARITY.
 FT ACT_SITE 239 239

CC YEAST 10:S75-S80(1994).
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE
 CC (BY SIMILARITY).
 CC -!- SIMILARITY: STRONG, TO SCOL.
 CC -!- SIMILARITY: SOME, TO B.SUBTILIS YPMO AND TO A.P.STUTZERI
 CC HYPOTHETICAL PROTEIN IN APT 3'REGION (AC P47206).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z35893; G536232; -;
 DR EMBL: X76078; G498753; -;
 DR PIR: S45880; S45880.
 DR SGD: L0001816; SC02.
 KW MITOCHONDRION; TRANSMEMBRANE; TRANSIT PEPTIDE.
 FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
 FT CHAIN ? 301 SC02 PROTEIN.
 FT TRANSMEM 82 98 POTENTIAL.
 SQ SEQUENCE 301 AA; 34888 MW; 41A6B5B1 CRC32;

Query Match 66.7%; Score 44; DB 1; Length 301;
 Best Local Similarity 44.4%; Pred. No. 5.32e+00;
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 186 SCDPARDTP 194
 |::|::|:
 QY 554 CEFQHOSP 562

RESULT 14
 ID YBEK_ECOLI STANDARD; PRT; 311 AA.
 AC P41409; P77738;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 33.8 KD PROTEIN IN LEUS-GLTL INTERGENIC REGION.
 GN YBEK.
 OS ESCHERICHIA COLI.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC ESCHERICHIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL SCIENCE 277:1453-1474(1997).
 CC [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA CHUNG E., ALLEN E., ARAUJO R., APARICIO A., DAVIS K., DUNCAN M.,
 RA FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O., LEW H., LIN D.,
 RA NAMATH A., OEFNER P., ROBERTS D., SCHRAMM S., DAVIS R.W.;
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC [3]
 RP SEQUENCE OF 1-123 FROM N.A.
 RC STRAIN-K12 / BK9WDG;
 RA LUM D., WALLACE B.J.;
 RL SUBMITTED (JUN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC [4]
 RP IDENTIFICATION.
 RA RUDD K.E.;
 RL UNPUBLISHED OBSERVATIONS (MAR-1995).
 CC -!- SIMILARITY: BELONGS TO THE IUNH FAMILY.
 CC

SQ SEQUENCE 313 AA: 33748 MW: 3BC554F1 CRC32;

Query Match 66.7%; Score 44; DB 1; Length 313;
Best Local Similarity 50.0%; Pred. No. 5.32e+00;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 10 CDPGHDDA 17

Qy 555 CEPGHQSP 562
|:|:|:|:

Search completed: Thu Jul 8 19:03:49 1999
Job time : 13 secs.

W P E R L E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:04:08 1999; MasPar time 9.24 Seconds

Tabular output not generated. 53.163 Million cell updates/sec

Title: >US-09-041-236-2
Description: (554-562) from US09041236.pep (28 of 45)
Perfect Score: 66

Sequence: 1 SCEPGHQSP 9

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprenbl9

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 20.003; Variance 22.069; scale 0.906

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	66	100.0	666	4	SEMAPHORIN L	1.38e-05
2	52	78.8	150	5	MARINER TRANSPOSASE (F	8.39e-02
3	50	75.8	836	5	TWEETY.	2.68e-01
4	49	74.2	35	14	UNIDENTIFIED ORF.	4.75e-01
5	49	74.2	738	11	PACKTOLUS.	4.75e-01
6	49	74.2	1014	11	CHICKEN EPH/ELK RECEPT	4.75e-01
7	46	69.7	1805	4	MYELOBLAST KIAA0257 (F	2.53e+00
8	45	68.2	386	1	386AA LONG HYPOTHETICA	4.36e+00
9	45	68.2	550	5	M89 PROTEIN.	4.36e+00
10	45	68.2	629	5	C29F3.2 PROTEIN.	4.36e+00
11	44	66.7	352	2	HYPOTHETICAL 37.0 KD P	7.46e+00
12	44	66.7	387	5	T27B7.3 PROTEIN.	7.46e+00
13	44	66.7	417	13	(TGCCA-BINDING PROTEIN	7.46e+00
14	44	66.7	510	13	(TGCCA-BINDING PROTEIN	7.46e+00
15	44	66.7	813	2	SENSOR-REGULATOR PROTE	7.46e+00
16	44	66.7	1388	14	FEELINE HERPESVIRUS TYP	7.46e+00
17	44	66.7	1584	5	F54F3.1 PROTEIN.	7.46e+00
18	43	65.2	306	11	RECEPTOR PROTEIN TYROS	1.26e+01
19	43	65.2	357	6	SORBITOL DEHYDROGENASE	1.26e+01
20	43	65.2	592	5	HAEMOCYANIN 1 (FRAGMEN	1.26e+01

21	43	65.2	669	4	075441	LATENT TRANSFORMING GR	1.26e+01
22	43	65.2	974	5	P91658	FURROWED.	1.26e+01
23	43	65.2	1372	5	P91526	SIMILARITY TO MULTIPLE	1.26e+01
24	43	65.2	1378	11	O62190	RECEPTOR PROTEIN TYROS	1.26e+01
25	43	65.2	1378	11	O62555	RECEPTOR PROTEIN TYROS	1.26e+01
26	43	65.2	1511	4	075412	LATENT TRANSFORMING GR	1.26e+01
27	43	65.2	1587	4	O00508	LATENT TGF-BETA BINDIN	1.26e+01
28	43	65.2	3857	11	O88840	MUTANT FIBRILLIN-1.	1.26e+01
29	42	63.6	132	2	O05750	HYPOTHETICAL 14.2 KD P	2.13e+01
30	42	63.6	310	2	O85727	METHYLTRANSFERASE CNCJ	2.13e+01
31	42	63.6	373	14	O56949	REGULATORY PROTEIN E2.	2.13e+01
32	42	63.6	708	13	P87363	FIBRILLIN-1 (FRAGMENT)	2.13e+01
33	42	63.6	1124	10	O49318	PUTATIVE RECEPTOR PROT	2.13e+01
34	42	63.6	1310	14	O41976	TEGUMENT PROTEIN/FGARA	2.13e+01
35	42	63.6	1567	11	O35243	NOTCH4.	2.13e+01
36	42	63.6	1964	11	O35442	ANTIGEN CONTAINING EPI	2.13e+01
37	42	63.6	3345	5	O46074	COSMID 30B8.	2.13e+01
38	41	62.1	135	5	O44956	C34B2.2 PROTEIN.	3.54e+01
39	41	62.1	173	2	P95260	HYPOTHETICAL 18.5 KD P	3.54e+01
40	41	62.1	263	3	O42899	HYPOTHETICAL 30.2 KD P	3.54e+01
41	41	62.1	543	5	O27043	POLYMORPHIC IMMUNODOMI	3.54e+01
42	41	62.1	736	6	O28482	TMDC I PRECURSOR.	3.54e+01
43	41	62.1	1074	5	O24028	DACHSHUND VARIANT 1.	3.54e+01
44	41	62.1	1764	11	O35806	LATENT TGF-BETA BINDIN	3.54e+01
45	41	62.1	4123	4	O75851	WUSC:H_DJ0751H13.1 PR	3.54e+01

ALIGNMENTS

RESULT	1				
ID	075326	PRELIMINARY;	PRT;	666	AA.
AC	075326;				
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L.				
GN	SEMAL.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;				
OC	CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98389619.				
RA	LANG E C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;				
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA				
RT	viruses."				
RL	GENOMICS 51:340-350(1998).				
DR	EMBL; AF030698; G3523115; ..				
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;				

Query Match 100.0%; Score 66; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 1.38e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	586	SCEPGHQSP	594
QY	554	SCEPGHQSP	562

RESULT 2
ID Q18591 PRELIMINARY; PRT; 150 AA.

AC	Q18591;				
DT	01-JAN-1998	(TREMBLREL. 05, CREATED)			
DT	01-JAN-1998	(TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DT	01-AUG-1998	(TREMBLREL. 07, LAST ANNOTATION UPDATE)			
DE	MARINER TRANSPOSASE (FRAGMENT).				
OS	PLEBEIA FRONTALIS.				
OC	EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;				
OC	PTERYGOTA; HYMENOPTERA; APOCRITA; ACULEATA; APOIDEA; PLEBEIA.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TRANSPOSON-MARINER TRANSPOSON;				
RA	ROBERTSON H.M., SOTO-ADAMES F.N., WALDEN K.K.O., AVANCINI R.M.P.,				

```
RA LAMPE D.J.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U91386; G2564426; -.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 17566 MW; 70A60E18 CRC32;

Query Match 78.8%; Score 52; DB 5; Length 150;
Best Local Similarity 66.7%; Pred. No. 8.39e-02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 47 SCEPGQAP 55
QY 554 SCEPGHQP 562

RESULT 3
ID O61343 PRELIMINARY; PRT; 836 AA.
AC O61343;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TWEEY.
GN TTY.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTONS;
RX MEDLINE: 96129280.
RA DE COUET H.G., FONG K.S., WEEDS A.G., MCLAUGHLIN P.J., MIKLOS G.L.;
RT "Molecular and mutational analysis of a gelsolin-family member
RT encoded by the flightless I gene of Drosophila melanogaster.";
RL GENETICS 141:1049-1059(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTONS;
RX MEDLINE: 97289742.
RA MIKLOS G.L., YAMAMOTO M., BURNS R.G., MALESZKA R.;
RT "An essential cell division gene of Drosophila, absent from
RT Saccharomyces, encodes an unusual protein with tubulin-like and
RT myosin-like peptide motifs.";
RL PROC. NATL. ACAD. SCI. U.S.A. 94:5189-5194(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTONS;
RA MALESZKA R., DE COUET H.G., MIKLOS G.L.G.;
RL PROC. NATL. ACAD. SCI. U.S.A. 0:0-0(1998).
DR EMBL; AF01777; G3004653; -.
SQ SEQUENCE 836 AA; 92838 MW; FE221871 CRC32;

Query Match 75.8%; Score 50; DB 5; Length 836;
Best Local Similarity 75.0%; Pred. No. 2.68e-01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 305 CEPGHTNP 312
QY 555 CEPGHQP 562

RESULT 4
ID Q98155 PRELIMINARY; PRT; 35 AA.
AC Q98155;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE UNIDENTIFIED ORF.
OS KAPOSI'S SARCOMA-ASSOCIATED HERPES-LIKE VIRUS.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE; RHADINOVIRUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=TUMOR;
RX MEDLINE: 96270595.
RA ZHONG W., WANG H., HERNDIER B., GANEM D.;
RT "Restricted expression of Kaposi sarcoma-associated herpesvirus
RT (human herpesvirus 8) genes in Kaposi sarcoma.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:6641-6646(1996).
DR EMBL; U69522; G1519300; -.
SQ SEQUENCE 35 AA; 3831 MW; B99223B3 CRC32;

Query Match 74.2%; Score 49; DB 14; Length 35;
Best Local Similarity 62.5%; Pred. No. 4.75e-01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 26 COPGHSS 33
QY 555 CEPGHQP 562

RESULT 5
ID O88424 PRELIMINARY; PRT; 738 AA.
AC O88424;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PACTOLUS.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH(S); TISSUE=BONE MARROW;
RX MEDLINE: 98204856.
RA CHEN Y., GARISON S., WEIS J.J., WEIS J.H.;
RT "Identification of pactolus, an integrin beta subunit-like
RT cell-surface protein preferentially expressed by cells of the bone
RT marrow.";
RL J. BIOL. CHEM. 273:8711-8718(1998).
CC -!- FUNCTION: INTEGRINS ARE A LARGE FAMILY OF CELL SURFACE
CC GLYCOPROTEINS THAT MEDIATE CELL TO CELL & CELL TO MATRIX
CC ADHESION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: WITH OTHER BETA CHAINS FROM THE INTEGRIN FAMILY OF
CC CELL-SURFACE RECEPTOR.
DR EMBL; AF051367; G3287491; -.
DR PROSITE; PS00243; INTEGRIN_BETA; 3.
KW INTEGRIN; CELL ADHESION; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT.
SQ SEQUENCE 738 AA; 81605 MW; BAD41DD5 CRC32;

Query Match 74.2%; Score 49; DB 11; Length 738;
Best Local Similarity 62.5%; Pred. No. 4.75e-01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 575 CDPGYQPP 582
QY 555 CEPGHQP 562

RESULT 6
ID O08644 PRELIMINARY; PRT; 1014 AA.
AC O08644;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CHICKEN EPH/ELK RECEPTOR-LIKE PROTEIN (MEP).
GN CEKL OR MEP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C X 129 F2; TISSUE=THYMUS;
RX MEDLINE: 96358627.
```

RA GURNIAK C.B., BERG L.J.;
RT "A new member of the Eph family of receptors that lacks protein
RT tyrosine kinase activity.";
RL ONCOGENE 13:777-786(1996).
DR EMBL: L77867; G191183; -.
DR MGD: MGI:108444; CEKL.
DR PFAM: PF00041; fn3; 2.
DR PFAM: PF00069; pkinase; 1.
DR PFAM: PF00536; SAM; 1.
SQ SEQUENCE 1014 AA; 110076 MW; F3723C39 CRC32;
Query Match 74.2%; Score 49; DB 11; Length 1014;
Best Local Similarity 62.5%; Pred. No. 4.75e-01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 296 CDPGHQPA 303
QY 555 CBPGHOSP 562
I:|||||:
I:|||||:
RESULT 7
ID Q92545 PRELIMINARY; PRT: 1805 AA.
AC Q92545;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MYELOBLAST KIAA0257 (FRAGMENT).
GN KIAA0257.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BONE MARROW;
RX MEDLINE; 97191544.
RA NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASI Y., OHARA O.,
RA TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA RES. 3:321-329(1996).
DR EMBL: D87446; D101407; -.
FT NON_TER 1
SQ SEQUENCE 1805 AA; 197588 MW; 9BDE8B0B CRC32;
Query Match 69.7%; Score 46; DB 4; Length 1805;
Best Local Similarity 83.3%; Pred. No. 2.53e-00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 1109 SCDPGH 1114
QY 554 SCEPGH 559
I:||||
I:||||
RESULT 8
ID Q59033 PRELIMINARY; PRT: 386 AA.
AC Q59033;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE 386AA LONG HYPOTHETICAL SERINE AMINOTRANSFERASE.
GN PH1308.
OS PYROCOCUS HORIKOSHII.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCOCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE K., KOSUGI H., HOSOIYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,

RA KIKUCHI H.;
RT "Complete Sequence and Gene Organization of the Genome of a
RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 5:55-76(1998).
DR EMBL: AP000006; D1031356; -.
KW TRANSFERASE; AMINOTRANSFERASE.
SQ SEQUENCE 386 AA; 43487 MW; 2D821345 CRC32;
Query Match 68.2%; Score 45; DB 1; Length 386;
Best Local Similarity 85.7%; Pred. No. 4.36e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 307 EPGHESP 313
QY 556 EPGHOSP 562
I:|||||:
I:|||||:
RESULT 9
ID Q94130 PRELIMINARY; PRT: 550 AA.
AC Q94130;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE M89 PROTEIN.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-N2;
RX MEDLINE; 96291396.
RA PORTER J.A., EKKER S.C., PARK W.J., YOUNG K.E., CHEN C.H., MA Y.,
RA WOODS A.S., COTTER R.J., KOONIN E.V., BEACHY P.A.;
RT "Hedgehog patterning activity: role of a lipophilic modification
RT mediated by the carboxy-terminal autoprocessing domain.";
RL CELL 86:21-34(1996).
DR EMBL: U61237; G1532157; -.
DR PFAM: PF01079; Hint; 1.
SQ SEQUENCE 550 AA; 62100 MW; BCF348 CRC32;
Query Match 68.2%; Score 45; DB 5; Length 550;
Best Local Similarity 66.7%; Pred. No. 4.36e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 99 TCEPGFQSS 107
QY 554 SCEPGHOSP 562
I:|||||:
I:|||||:
RESULT 10
ID Q45273 PRELIMINARY; PRT: 629 AA.
AC Q45273;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE C29F3.2 PROTEIN.
GN C29F3.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA MATTHEWS L.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KESHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WORLDMAN P.; III of C.
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL NATURE 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA WHITE S.;
 RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: Z81043; E1344265; -.
 DR EMBL: AL023813; E1344265; JOINED.
 DR EMBL: AL023813; E1347656; -.
 DR EMBL: Z81043; E1347656; JOINED.
 SQ SEQUENCE 629 AA; 71349 MW; F266C02D CRC32;
 Query Match 68.2%; Score 45; DB 5; Length 629;
 Best Local Similarity 66.7%; Pred. No. 4.36e+00;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 178 TCPEGFQSS 186
 Qy 554 SCEPGHQP 562

RESULT 11
 ID O07211 PRELIMINARY; PRT; 352 AA.
 AC O07211;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 37.0 KD PROTEIN.
 GN MTC105A6.33C.
 OS MYCOBACTERIUM TUBERCULOSIS.
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
 OC ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV.
 RA GENTLES S., CHURCHER C.M.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA BARRELL B.G., RAJANDREAM M.A., PARKHILL J.;
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE; 96181548.
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RT "An integrated map of the genome of the tubercle bacillus,
 RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
 RT leorae.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
 DR EMBL: Z96072; E321105; -.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 352 AA; 36957 MW; B3CEB015 CRC32;

Query Match 66.7%; Score 44; DB 2; Length 352;
 Best Local Similarity 83.3%; Pred. No. 7.46e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 300 CEPGHR 305
 Qy 555 CEPGHQ 560
 RESULT 12
 ID O16661 PRELIMINARY; PRT; 387 AA.

AC O16661;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE T27B7.3 PROTEIN.
 GN T27B7.3.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WORLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA DANTE M., WAMSLEY P., O'BRIEN D.;
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF016675; G2315695; -.
 DR PFAM: PF00105; zf-C4; 1.
 SQ SEQUENCE 387 AA; 45642 MW; EABEEF93 CRC32;
 Query Match 66.7%; Score 44; DB 5; Length 387;
 Best Local Similarity 50.0%; Pred. No. 7.46e+00;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Db 103 CDPHSAP 110
 Qy 555 CEPGHQP 562
 RESULT 13
 ID Q30931 PRELIMINARY; PRT; 417 AA.
 AC Q30931;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE (TGGCA-BINDING PROTEIN) (FRAGMENT).
 GN CNFI-C.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94238700.
 RA KRUSE U., SIPPEL A.E.;
 RT "The genes for transcription factor nuclear factor I give rise to
 RT corresponding splice variants between vertebrate species.";
 RL J. MOL. BIOL. 238:860-865(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94298952.
 RA KRUSE U., SIPPEL A.E.;
 RT "Transcription factor nuclear factor I proteins form stable homo- and
 RT heterodimers.";

```
RL FEBS LETT. 348:46-50(1994).
CC -!- FUNCTION: CTF/NF-I RECOGNIZE AND BIND THE PALINDROMIC SEQUENCE
CC 5'-TGCCANNNTGCCA-3'PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2.
CC THESE PROTEINS ARE INDIVIDUALLY CAPABLE OF ACTIVATING
CC TRANSCRIPTION AND REPLICATION (BY SIMILARITY).
CC -!- SUBUNIT: BINDS DNA AS AN HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: HIGH, TO OTHER CTF/NF-I PROTEINS IN VERTEBRATES.
CC EMBL; X68157; G63682; -.
DR PROSITE; PS00349; CTF_NFI; 1.
DR PFAM; PF00859; CTF_NFI; 1.
KW TRANSCRIPTION REGULATION; DNA REPLICATION; DNA-BINDING; ACTIVATOR;
KW NUCLEAR PROTEIN; ALTERNATIVE SPLICING.
FT NON_TER 1
SQ SEQUENCE 417 AA; 46743 MW; 43EB7037 CRC32;

Query Match 66.7%; Score 44; DB 13; Length 417;
Best Local Similarity 44.4%; Pred. No. 7.46e+00;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 402 ACDPNQOP 410
QY 554 SCEPGHQP 562
:|:|:|

RESULT 14 PRELIMINARY; PRT; 510 AA.
ID Q90930
AC Q90930;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE (TGCA-BINDING PROTEIN).
GN CNFI-C.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94238700.
RA KRUSE U., SIPPEL A.E.;
RT "The genes for transcription factor nuclear factor I give rise to
RT corresponding splice variants between vertebrate species.";
RL J. MOL. BIOL. 238:860-865(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94298952.
RA KRUSE U., SIPPEL A.E.;
RT "Transcription factor nuclear factor I proteins form stable homo- and
RT heterodimers.";
RL FEBS LETT. 348:46-50(1994).
CC -!- FUNCTION: CTF/NF-I RECOGNIZE AND BIND THE PALINDROMIC SEQUENCE
CC 5'-TGCCANNNTGCCA-3'PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2.
CC THESE PROTEINS ARE INDIVIDUALLY CAPABLE OF ACTIVATING
CC TRANSCRIPTION AND REPLICATION (BY SIMILARITY).
CC -!- SUBUNIT: BINDS DNA AS AN HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: HIGH, TO OTHER CTF/NF-I PROTEINS IN VERTEBRATES.
CC EMBL; X68158; G63678; -.
DR PROSITE; PS00349; CTF_NFI; 1.
DR PFAM; PF00859; CTF_NFI; 1.
KW TRANSCRIPTION REGULATION; DNA REPLICATION; DNA-BINDING; ACTIVATOR;
KW NUCLEAR PROTEIN; ALTERNATIVE SPLICING.
SQ SEQUENCE 510 AA; 56295 MW; B442A23F CRC32;

Query Match 66.7%; Score 44; DB 13; Length 510;
Best Local Similarity 44.4%; Pred. No. 7.46e+00;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 413 ACDPNQOP 421
QY 554 SCEPGHQP 562
:|:|:|
```

```
RESULT 15 PRELIMINARY; PRT; 813 AA.
ID P77032
AC P77032;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE SENSOR-REGULATOR PROTEIN BARA (FRAGMENT).
GN BARA.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE; 97349980.
RA YAMAMOTO Y., AIBA H., BABA T., HAYASHI K., INADA T., ISONO K.,
RA ITOH T., KIMURA S., KITAGAWA M., MAKINO K., MIKI T., MITSUHASHI N.,
RA MIZOBUCHI K., MORI H., NAKADE S., NAKAMURA Y., NASHIMOTO H.,
RA OSHIMA T., OYAMA S., SAITO N., SAMPEI G., SATOH Y., SIVASUNDARAM S.,
RA TAGAMI H., TAKAHASHI H., TAKEDA J., TAKEMOTO K., UEHARA K., WADA C.,
RA YAMAGATA S., HORIUCHI T.;
RT "Construction of a contiguous 874 kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min region on the linkage map
RT and analysis of its sequence features.";
RL DNA RES. 4:91-113(1997).
DR EMBL; D90894; D1017304; -.
DR PFAM; PF00072; response_reg; 1.
DR PFAM; PF00512; signal; 1.
DR PFAM; PF00672; DUF5; 1.
FT NON_TER 1
SQ SEQUENCE 813 AA; 90471 MW; 85368AC2 CRC32;

Query Match 66.7%; Score 44; DB 2; Length 813;
Best Local Similarity 57.1%; Pred. No. 7.46e+00;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 592 CDSGHQA 598
QY 555 CEPGHQS 561
:|:|:|
```

Search completed: Thu Jul 8 19:04:23 1999
Job time : 15 secs.

THIS PAGE BLANK (USPTO)

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:07:38 1999; MasPar time 11.04 Seconds
Tabular output not generated.
23.121 Million cell updates/sec

Title: >US-09-041-236-2
Description: (572-583) from US09041236.pap (29 of 45)
Perfect Score: 97
Sequence: 1 TAOQYGHYFCEA 12

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 18.206; Variance 72.482; scale 0.251

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	62	63.9	116	30	W40416	5.29e+01
2	58	59.8	402	7	Human OP-2-pp.	1.16e+02
3	58	59.8	642	9	Cephalosporin C #1.	1.16e+02
4	57	58.8	139	4	Osteogenic protein fo	1.40e+02
5	57	58.8	377	16	R48593	1.40e+02
6	57	58.8	377	19	G-protein coupled hum	1.40e+02
7	57	58.8	466	5	Adrenalin receptor su	1.40e+02
8	56	57.7	97	8	Human osteogenic prot	1.70e+02
9	56	57.7	109	30	8019 vK antibody.	1.70e+02
10	56	57.7	133	39	W95459	1.70e+02
11	56	57.7	133	8	Mature species of hum	1.70e+02
12	56	57.7	136	39	Mature human OP2, hOP	1.70e+02
13	56	57.7	137	8	Murine osteogenic pro	1.70e+02
14	56	57.7	139	39	Mature human osteogen	1.70e+02
15	56	57.7	139	7	Human mature OP-2.	1.70e+02
16	56	57.7	139	10	Human OP-2.	1.70e+02

17	56	57.7	139	9	R46735	Mouse osteogenic prot	1.70e+02
18	56	57.7	139	7	R33524	Morphogen mop-2.	1.70e+02
19	56	57.7	139	7	R33401	Mouse mature OP-2.	1.70e+02
20	56	57.7	139	8	R53393	Human osteogenic prot	1.70e+02
21	56	57.7	139	7	R33923	Morphogen hOP-2.	1.70e+02
22	56	57.7	139	31	W54059	Bone morphogenetic pr	1.70e+02
23	56	57.7	139	9	R46734	Human mature osteogen	1.70e+02
24	56	57.7	139	33	W63004	Human OP2 protein use	1.70e+02
25	56	57.7	160	39	W95460	Mature human OP2, hOP	1.70e+02
26	56	57.7	160	8	R53394	Mature species of hum	1.70e+02
27	56	57.7	281	7	R36737	Human BMP-8.	1.70e+02
28	56	57.7	399	9	R47253	mOP2.	1.70e+02
29	56	57.7	399	39	W89591	Murine osteogenic pro	1.70e+02
30	56	57.7	399	7	R33411	Murine pro-OP-2-pp.	1.70e+02
31	56	57.7	399	9	R47293	mOP2.	1.70e+02
32	56	57.7	399	25	W23144	Bone morphogenic prot	1.70e+02
33	56	57.7	399	19	W00239	Mouse osteogenic prot	1.70e+02
34	56	57.7	399	10	R50201	Murine OP-2.	1.70e+02
35	56	57.7	399	1	R23846	Prepro form of human	1.70e+02
36	56	57.7	399	30	W53182	Mouse osteogenic prot	1.70e+02
37	56	57.7	402	39	W95456	Human osteogenic prot	1.70e+02
38	56	57.7	402	39	W89592	Human osteogenic prot	1.70e+02
39	56	57.7	402	10	R57973	Human OP-2.	1.70e+02
40	56	57.7	402	8	R44759	Human osteogenic prot	1.70e+02
41	56	57.7	402	5	R27291	Human osteogenic prot	1.70e+02
42	56	57.7	402	23	W16366	Human osteogenic prot	1.70e+02
43	56	57.7	402	10	R50238	Human hippocampal ost	1.70e+02
44	56	57.7	402	26	W36855	Human OP-2.	1.70e+02
45	56	57.7	402	9	R47292	Full length sequence	1.70e+02
					hOP2.		

ALIGNMENTS

RESULT 1

ID W40416 standard; protein: 116 AA.
AC W40416:
DT 23-JUL-1998 (first entry)
DE E. coli flavodoxin protein fragment.
KW Nitric oxide synthase; endothelial; ENOS; INOS; inducible; NNOS;
KW neuronal; calmodulin; brain damage; shock; autoimmune disease;
KW inflammatory condition; multiple sclerosis; diabetes; dementia;
KW dysplasia; cancer; infectious disease; cytotoxic; hypertension;
KW atherosclerosis; asthma; detection; screening; flavodoxin.
OS Escherichia coli.
PN W09802555-A1.
PS 22-JAN-1998.
PF 10-JUL-1997: U12568.
PR 12-JUL-1996: US-679006.
PA (SALE/) SALERNO J C.
PI Salerno JC;
DR WPI; 98-110601/10.
PT New modulators of nitric oxide synthase - used for treating e.g.
PT toxic shock, auto-immune disease, inflammatory disease, diabetes,
PT hypertension, infections or cancer
PS Example 1; Fig 1; 78pp; English.
CC This sequence represents a E. coli flavodoxin fragment which is used to
CC describe a novel method that identifies an agent which inhibits nitric
CC oxide synthase (NOS) by blocking calmodulin (CAM) activation of the NOS.
CC Such agents which modulate NOS activity can be used to treat a disease or
CC condition associated with nitric oxide production. Agents which inhibit
CC neuronal NOS (NNOS) can be used to prevent brain damage in conditions
CC involving cerebral ischaemia or reperfusion injury, such as head trauma.
CC Agents which decrease the activity of inducible NOS (INOS) can be used to
CC treat a condition modulated by production of NO by INOS, such as septic
CC shock, toxic shock, autoimmune disease such as rheumatoid arthritis,
CC inflammatory conditions such as inflammatory bowel disease, multiple
CC sclerosis, diabetes, or to combat dementia, immune system destruction,
CC and/or physical deterioration in individuals infected with the AIDS
CC virus. Agents which activate INOS can be used to treat disease relating
CC to dysplasia, cancer, or infectious disease. Activation of INOS can
CC produce cytotoxic levels of NO which would aid in the elimination of
CC dysplastic or cancerous tissue, or aid in the control of infectious
CC agents such as viruses, microbes, or other parasites. Agents which

CC increase the activity of endothelial NOS (ENOS) can be used to treat a
CC condition modulated by production of NO by ENOS such as hypertension,
CC atherosclerosis or acute asthma. An agent which activates ENOS or NNOS in
CC the corpus cavernosa can be used for treating male erectile dysfunction.
CC The products and methods can also be used for detection and drug
CC screening. 116 AA;
SQ Sequence 116 AA;

Query Match 63.9%; Score 62; DB 30; Length 116;
Best Local Similarity 50.0%; Pred. No. 5.29e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 91 edyaeyfeda 100
::: |::|
QY 574 QOQYGHVFCEA 583

RESULT 2
ID R33410 standard; Protein: 402 AA.
AC R33410;
DT 15-JUL-1993 (first entry)
DE Human Op-2-PP.
KW morphogenic; osteogenic protein; developmental cascade; hOP-2;
KW inflammation; anti-inflammatory; Transforming Growth Factor;
KW TGF-beta super-family; hippocampus.
OS Homo sapiens.
FH Key Location/Qualifiers
FT protein 264..402
FT /note="contains conserved 7 cysteine skeleton"
PN W09304692-A.
PD 18-MAR-1993.
PF 28-AUG-1992; U07358.
PR 30-AUG-1991; US-752764.
PR 30-AUG-1991; US-752861.
PR 30-AUG-1991; US-753059.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Cohen CM, Kuberasampath T, Oppermann H, Ozkaynak E;
PI Pang RHL, Rueger DC, Smart JE;
DR WPI: 93-100652/12.
DR N-PSDB: Q38735.
PT Morphogen-induced modulation of inflammatory response - and
PT resulting tissue damage, e.g. in autoimmune diseases, diabetes,
PT asthma, ischemia reperfusion injury, etc.
PS Claim 26; Page 121-123; 165pp; English.
CC Human osteogenic protein (OP)-2 is a preferred morphogen for use in
CC treating tissue damage in e.g. inflammatory disease, autoimmune
CC disease, arthritis, psoriasis, dermatitis, diabetes and emphysema.
CC Proteins having at least 70% homology with OP-2 amino acid
CC sequences can also be used. See R33400 for mature hOP-2.
SQ Sequence 402 AA;

Query Match 59.8%; Score 58; DB 7; Length 402;
Best Local Similarity 41.7%; Pred. No. 1.16e+02;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 321 apqaysayyceg 332
::: |::|
QY 572 TAQYGHVFCEA 583

RESULT 3
ID R49826 standard; Protein: 642 AA.
AC R49826;
DT 30-SEP-1994 (first entry)
DE Cephalosporin C #1.
KW Cephalosporin C; A. crisogenum; biosynthesis; fermentation.
OS Acremonium chrysogenum.
PN J06038763-A.
PD 15-FEB-1994.
PF 12-MAR-1993; 077367.
PR 13-MAR-1992; JP-088418.
PA (ASAH) ASAH CHEM IND CO LTD.
DR WPI: 94-094831/12.

DR N-PSDB: Q44446.
PT A DNA fragment contg. a gene used in the bio-synthesis of
PT cephalosporin C - used to produce recombinants with improved
PT biosynthesis of cephalosporin C
PS Claim 3; Page 9-12; 19pp; Japanese.
CC The sequences given in R49826-27 are encoded the cephalosporin C gene
CC which was isolated from A. crisogenum. The gene fragments encoding
CC these proteins may be used in the biosynthesis of cephalosporin C and
CC in the production of A. crisogenum strains with improved cephalosporin
CC C fermentation capabilities.
SQ Sequence 642 AA;

Query Match 59.8%; Score 58; DB 9; Length 642;
Best Local Similarity 100.0%; Pred. No. 1.16e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 93 qyghyf 98
|::|::|
QY 575 QYGHYF 580

RESULT 4
ID R23187 standard; protein; 139 AA.
AC R23187;
DT 01-JAN-1980 (first entry)
DE Osteogenic protein for dispersal in implantable osteogenic device.
KW Osteogenic; implant; endochondral; bone; Op1.
OS mammalian.
PN W09118558-A.
PD 12-DEC-1991.
PF 22-MAY-1991; U03603.
PR 29-MAY-1990; US-529852.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Kuberasampath T, Berlowitz TL;
DR WPI: 92-007153/01.
PT Osteogenic device for implantation - comprising osteogenic
PT protein dispersed in porous matrix of collagen and
PT glycosaminoglycan
PS Claim 8; Page 34; 39pp; English.
CC The invention relates to an implantable device for inducing
CC endochondral bone formation in a shape conforming to the shape of the
CC device. The device comprises an osteogenic protein dispersed in a
CC porous matrix which is a polymer of collagen and glycosaminoglycan
CC crosslinked to an Mc value of 800 to 60,000. The osteogenic protein
CC comprises a pair of subunits constituting a stable dimer under
CC oxidising conditions. The present sequence (designated Op1), when
CC dimerised to form a homodimer or a heterodimer with certain other
CC protein sequences, can induce endochondral bone formation and hence
CC is preferred as one of the pair of subunits.
SQ Sequence 139 AA;

Query Match 58.8%; Score 57; DB 4; Length 139;
Best Local Similarity 33.3%; Pred. No. 1.40e+02;
Matches 4; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 58 apegaryyceg 69
::: |::|
QY 572 TAQYGHVFCEA 583

RESULT 5
ID R48693 standard; Protein; 377 AA.
AC R48693;
DT 05-JUN-1996 (first entry)
DE G-protein coupled human beta-1 adrenergic receptor protein.
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW psychotic disorder; schizophrenia; dopamine; CAMP; adenosine; thrombin;
KW muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;
KW rhodopsin; opsin; odorant; cytomagalovirus.
OS Homo sapiens.
PN W09405695-A1.
PD 17-MAR-1994.
PF 09-SEP-1993; U08528.

PR 10-SEP-1992; US-943236.
PA (UUNY) UNIV NEW YORK STATE.
PI Murphy RB, Schuster DI;
DR WPI; 94-101120/12.
PT Polypeptides of G-coupled receptor proteins (GPRs) - useful for
PS binding GPR ligands or modulating GPR binding
PS Disclosure: Page 65-66; 160pp; English.
CC Proteins R4865-R48758 represent a range of G-protein coupled receptor
CC proteins selected from CAMP, adenosine, muscarinic acetylcholine,
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC odorant, cytomagaloviral and other G-protein coupled receptors. The
CC receptor proteins were used to design polypeptides, pref. based on the
CC transmembrane domains, for use in G-protein coupled receptor ligand
CC binding assays. The polypeptide fragments retain biological activity
CC such as binding a GPR ligand or modulating GPR ligand binding to a GPR
CC (see R48759-R48758, R50569-R50807 and R89189-R89195 for examples of
CC polypeptide fragments). The polypeptide fragments can be used in
CC compositions for treating subjects suffering from a pathology related to
CC a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
SQ Sequence 377 AA;

Query Match 58.8%; Score 57; DB 16; Length 377;
Best Local Similarity 62.5%; Pred. No. 1.40e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 67 eygsffce 74

:||:||||

Qy 575 QYGHYFCE 582

RESULT 6
ID W02665 standard; peptide; 377 AA.
AC W02665;

DE 12-NOV-1996 (first entry)
DE G-protein coupled human beta-1 adrenergic receptor variant #1.
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW schizophrenia; dopamine; CAMP; adenosine; thrombin; adrenergic; opsin;
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW odorant; cytomagalovirus; serotonergic.
OS Homo sapiens.
PN US5508384-A.
PD 16-APR-1996.
PE 10-SEP-1992; 943236.
PR 10-SEP-1992; US-943236.
PR 09-SEP-1993; US-118270.
PA (UUNY) UNIV NEW YORK STATE.
PI Murphy RB, Schuster DI;
DR WPI; 96-208785/21.
PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.
PT for treating schizophrenia
PS Disclosure: Column 57-60; 184pp; English.
CC Proteins W02657-W02720 represent a range of G-protein coupled receptor
CC (GPR) proteins selected from CAMP, adenosine, muscarinic acetylcholine,
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC odorant, cytomagaloviral and other GPR proteins. The receptor proteins
CC were used to design polypeptides, pref. based on the transmembrane
CC domains, for use in G-protein coupled receptor ligand binding assays.
CC The polypeptide fragments retain biological activity such as binding a
CC GPR ligand or modulating GPR ligand binding to a GPR (see W02747-W02999
CC for examples of polypeptide fragments). The polypeptide fragments can
CC be used in compositions for treating subjects suffering from a pathology
CC related to a GPR abnormality e.g. a psychotic disorder such as
CC schizophrenia.
SQ Sequence 377 AA;

Query Match 58.8%; Score 57; DB 19; Length 377;
Best Local Similarity 62.5%; Pred. No. 1.40e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 67 eygsffce 74

:||:||||

Qy 575 QYGHYFCE 582

RESULT 7
ID R26505 standard; Protein; 466 AA.
AC R26505;
DE 12-FEB-1993 (first entry)
DE Adrenalin receptor subtype beta-1.
KW Drugs; dibutamine; prenalterol; acebutolol; metoprolol; atenolol;
KW practolol.
OS Rattus rattus.
PN J04211374-A.
PD 03-AUG-1992.
PE 15-FEB-1991; 044479.
PR 27-FEB-1990; JP-048794.
PA (SUMU) SUMITOMO SEIYAKU KK.
DR WPI; 92-304939/37.
DR N-PSDB; 027816.
PT DNA encoding adrenalin receptor sub-type beta-1 - obtd.
PT by cloning rat genome DNA library and rat cDNA library with probe
PT derived from human adrenalin
PS Disclosure; Page 5; 8pp; Japanese.
CC The sequence is useful in studies of drugs acting on adrenalin
CC receptor subclass beta 1, e.g. dobutamine or prenalterol, or drugs
CC blocking the receptor, e.g. acebutolol, metoprolol, atenolol or
CC practolol.
SQ Sequence 466 AA;

Query Match 58.8%; Score 57; DB 5; Length 466;
Best Local Similarity 62.5%; Pred. No. 1.40e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 125 eygsffce 132

:||:||||

Qy 575 QYGHYFCE 582

RESULT 8
ID R53390 standard; Protein; 97 AA.
AC R53390;
DE 06-JUN-1994 (first entry)
DE Human osteogenic protein Op2 fragment.
KW Osteogenic protein; bone; cartilage; matrix; osteoarthritis;
KW repair; vascularisation; mineralisation; differentiation.
OS Homo sapiens.
PN US5266683-A.
PD 30-NOV-1993.
PE 08-APR-1988; 179406.
PR 08-APR-1988; US-179406.
PR 15-AUG-1988; US-232630.
PR 23-FEB-1989; US-315342.
PR 17-OCT-1989; US-422613.
PR 17-OCT-1989; US-422699.
PR 22-FEB-1990; US-483913.
PR 20-AUG-1990; US-569920.
PR 07-SEP-1990; US-579865.
PR 18-OCT-1990; US-599543.
PR 18-OCT-1990; US-600024.
PA 04-DEC-1990; US-621849.
PA 04-DEC-1990; US-621988.
PA 22-FEB-1991; US-660162.
PA 20-DEC-1991; US-810560.
PA 28-JAN-1992; US-827052.
PA 21-FEB-1992; US-841646.
PI Kuberasanpath T, Oppermann H, Ozkaynak E, Pang RHL;
PI Rueger DC;
DR WPI; 93-395405/49.
DR N-PSDB; 053155.
PT New pure mammalian osteogenic proteins - induce cartilage and
PT endochondral bone formation when in association with a matrix
PS Claim 21: Columns 145-148; 128pp; English.
CC The osteogenic protein when in association with a matrix can induce
CC at the locus of an implant the full development cascade of
CC endochondral bone formation including vascularisation,
CC mineralisation and bone marrow differentiation. The osteogenic

CC protein can also be used to repair both bone and cartilage in the
CC treatment of osteoarthritis. This sequence contains a 6 cysteine
CC skeleton conserved among osteogenic proteins in their active
CC regions.
SQ Sequence 97 AA;

Query Match 57.7%; Score 56; DB 8; Length 97;
Best Local Similarity 41.7%; Pred. No. 1.70e+02;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 16 apqgysayyceg 27

:::|:|:|:

QY 572 TAQQYGHYFCEA 583

RESULT 9

ID W46483 standard; protein; 109 AA.
AC W46483;
DT 24-JUN-1998 (first entry)
DE 8019 VK antibody.
KW Diagnosis: photoaffinity compound; therapeutic; antibody; SIC5; 8019;
KW antigen; immuno-conjugate.
OS Homo sapiens.
PN W09803870-A1.
PD 29-JAN-1998.
PF 22-JUL-1997; U12223.
PR 23-JUL-1996; US-681432.
PA (KENT) UNIV KENTUCKY RES FOUND.
PI Haley BE, Kohler H, Pavlinkova G, Rajogopalan K;
DR WPI; 98-120935/11.
PT Conjugates formed by site-specific attachment of nucleic
PT photo-affinity agents to antibodies - useful as immunoassay reagents
PT for detecting antigen, also immuno-therapeutic complex also
PT containing cytotoxin for treating cancer, infection or genetic
PT disease.
PS Disclosure: Figure 12c; 75pp; English.
CC This is the amino acid sequence of 8019 VK antibody. The invention
CC relates to the site-specific photo-attachment of a nucleic
CC photoaffinity compound to an antibody which comprises reacting them
CC under conditions that promote attachment of the photoaffinity
CC compound to at least 1 nucleotide binding site in the antibody. Also
CC claimed are: photoaffinity compound-antibody conjugates produced this
CC way, and immuno-conjugates comprising cytotoxic or therapeutic agent
CC site-specifically attached to the photoaffinity compound in the
CC conjugate. When the conjugate is labelled it is used for detection or
CC quantitation of an antigen (Ag) by any standard immunoassay format,
CC while immuno-conjugates are used therapeutically, e.g. in cases of
CC tumours, infections or genetic diseases, or as imaging agents. The
CC photoaffinity compounds are attached under mild, physiological
CC conditions, particularly a single 2-5 minute photoactivation results in
CC almost 100% attachment (contrast conditions usually required to attach
CC molecules to antibodies.
SQ Sequence 109 AA;

Query Match 57.7%; Score 56; DB 30; Length 109;
Best Local Similarity 40.0%; Pred. No. 1.70e+02;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 80 pedfgsyfcq 89

:::|:|:|:

QY 573 AQQYGHYFCE 582

RESULT 10

ID W95459 standard; protein; 133 AA.
AC W95459;
DT 26-MAR-1999 (first entry)
DE Mature human OP2, hOP2-Arg (residues 270-402).
KW Cystic kidney disease; renal; therapeutic; osteogenic protein; OP;
KW bone morphogenic protein; BMP; growth factor-beta superfamily;
KW polycystic kidney disease; multicystic dysplastic kidney disease;
KW uraemic medullary cystic disease; human; OPI; truncated.
OS Homo sapiens.

PN W09850061-A1.
PD 12-NOV-1998.
PF 06-MAY-1998; U09268.
PR 07-MAY-1997; US-045909.
PA (BIOJ) BIOGEN INC.
PI GJorstrup P, Harris R;
DR WPI; 99-070084/06.
PT Treating cystic kidney disease - using renal therapeutic agents or
PT sequences encoding them, especially from the osteogenic protein/bone
PT morphogenic protein family
PS Claim 4; Page -; 67pp; English.
CC The invention relates to methods for treating cystic kidney diseases. The
CC method comprises administering an effective amount of a renal therapeutic
CC agent or a polynucleotide encoding the therapeutic agent. The therapeutic
CC agent is preferably a soluble or membrane bound polypeptide, e.g. a
CC member of the osteogenic protein/bone morphogenic protein (OP/BMP) family
CC within a transforming growth factor-beta superfamily of proteins. It is
CC especially one of the polypeptides hOP1, hOP1-PP, OPI-18Ser, OPS, OP7,
CC OPI-16Ser, OPI-16Leu, OPI-16Met, OPI-16Ala, OPI-16Val, MOPI, MOPI-PP,
CC hOP2, hOP2-PP, hOP2-Ala, hOP2-Pro, hOP2-Arg, or hOP2-Ser or their
CC biologically active homologues. The method is used to treat humans
CC having, or at risk of, cystic kidney disease, e.g. autosomal recessive
CC (infantile) polycystic disease, multicystic dysplastic kidney disease,
CC uraemic medullary cystic disease, and autosomal dominant polycystic
CC kidney disease. The present sequence represents possible mature human
CC osteogenic protein 2 (OP2) that can be used as a therapeutic agent in
CC the method of the invention.
CC Note: This sequence is not provided in the specification. It has been
CC derived from the human OP2 sequence (W95456) obtained from U.S. Patent
CC No. 5,266,683.
SQ Sequence 133 AA;

Query Match 57.7%; Score 56; DB 39; Length 133;
Best Local Similarity 41.7%; Pred. No. 1.70e+02;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 52 apqgysayyceg 63

:::|:|:|:

QY 572 TAQQYGHYFCEA 583

RESULT 11

ID R53391 standard; Protein; 133 AA.
AC R53391;
DT 06-JUN-1994 (first entry)
DE Mature species of human osteogenic protein OP2.
KW Osteogenic protein; Bone; cartilage; matrix; osteoarthritis;
KW repair; vascularisation; mineralisation; differentiation.
OS Homo sapiens.
PN US5266683-A.
PD 30-NOV-1993.
PF 08-APR-1988; 179406.
PR 08-APR-1988; US-179406.
PR 15-AUG-1988; US-232630.
PR 23-FEB-1989; US-315342.
PR 17-OCT-1989; US-422613.
PR 17-OCT-1989; US-422699.
PR 22-FEB-1990; US-483913.
PR 20-AUG-1990; US-569920.
PR 07-SEP-1990; US-579865.
PR 18-OCT-1990; US-599543.
PR 18-OCT-1990; US-600024.
PA 04-DEC-1990; US-621849.
PA 22-FEB-1991; US-660162.
PA 20-DEC-1991; US-810560.
PA 28-JAN-1992; US-827052.
PA 21-FEB-1992; US-841646.
PI Kuberasampath T, Oppermann H, Ozkaynak E, Pang RHL;
PI Rueger DC;
DR WPI; 93-395405/49.
DR N-PSDB; Q53155.
PT New pure mammalian osteogenic proteins - induce cartilage and

PT endochondral bone formation when in association with a matrix
 PS Claim 22: Columns 145-148; 128pp; English.
 CC The osteogenic protein when in association with a matrix can induce
 CC at the locus of an implant the full development cascade of
 CC endochondral bone formation including vascularisation,
 CC mineralisation and bone marrow differentiation. The osteogenic
 CC protein can also be used to repair both bone and cartilage in the
 CC treatment of osteoarthritis. This sequence is a mature species of
 CC the pre-pro form of the protein (R44759) designated hOP2-Arg and is
 CC thought to be active.
 SQ Sequence 133 AA;

Query Match 57.7%; Score 56; DB 8; Length 133;
 Best Local Similarity 41.7%; Pred. No. 1.70e+02;
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 52 apqgysayyceg 63
 QY 572 TAQYGHVFCEA 583
 :|:|:|:|:

RESULT 12
 ID W95458 standard; protein; 136 AA.

AC W95458;
 DT 26-MAR-1999 (first entry)
 DE Mature human OP2, hOP2-Pro (residues 267-402).
 KW Cystic kidney disease; renal; therapeutic; osteogenic protein; OP;
 KW bone morphogenic protein; BMP; growth factor-beta superfamily;
 KW polycystic kidney disease; multicystic dysplastic kidney disease;
 KW uraemic medullary cystic disease; human; OPI; truncated.
 OS Homo sapiens.
 PN WQ9850061-A1.
 PD 12-NOV-1998.
 PF 06-MAY-1998; U09268.
 PR 07-MAY-1997; US-045909.
 PA (BIOJ) BIOGEN INC.
 PI Gjostrup P, Harris R;
 DR WPI: 99-070084/06.
 PT Treating cystic kidney disease - using renal therapeutic agents or
 PT sequences encoding them, especially from the osteogenic protein/bone
 PT morphogenic protein family
 PS Claim 4: Page -: 67pp; English.

CC The invention relates to methods for treating cystic kidney diseases. The
 CC method comprises administering an effective amount of a renal therapeutic
 CC agent or a polynucleotide encoding the therapeutic agent. The therapeutic
 CC agent is preferably a soluble or membrane bound polypeptide, e.g. a
 CC member of the osteogenic protein/bone morphogenic protein (OP/BMP) family
 CC within a transforming growth factor-beta superfamily of proteins. It is
 CC especially one of the polypeptides hOP1, hOP1-PP, OP1-18Ser, OP5, OP7,
 CC OP1-16Ser, OP1-16Ieu, OP1-16Met, OP1-16Ala, OP1-16Val, mOP1, mOP1-PP,
 CC hOP2, hOP2-PP, hOP2-Ala, hOP2-Pro, hOP2-Arg, or hOP2-Ser or their
 CC biologically active homologues. The method is used to treat humans
 CC having, or at risk of, cystic kidney disease, e.g. autosomal recessive
 CC (infantile) polycystic disease, multicystic dysplastic kidney disease,
 CC uraemic medullary cystic disease, and autosomal dominant polycystic
 CC kidney disease. The present sequence represents possible mature human
 CC osteogenic protein 2 (OP2) that can be used as a therapeutic agent in
 CC the method of the invention.
 CC Note: This sequence is not provided in the specification. It has been
 CC derived from the human OP2 sequence (W95456) obtained from U.S. Patent
 CC No. 5,266,683.
 SQ Sequence 136 AA;

Query Match 57.7%; Score 56; DB 39; Length 136;
 Best Local Similarity 41.7%; Pred. No. 1.70e+02;
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 55 apqgysayyceg 66
 QY 572 TAQYGHVFCEA 583
 :|:|:|:|:

RESULT 13

ID R53388 standard; Protein; 137 AA.
 AC R53388;
 DT 06-JUN-1994 (first entry)
 DE Murine osteogenic protein OP2.
 KW Osteogenic protein; bone; cartilage; matrix; osteoarthritis;
 KW repair; vascularisation; mineralisation; differentiation.
 OS Mus musculus.
 PN US5266683-A.
 PD 30-NOV-1993.
 PF 08-APR-1988; 179406.
 PR 08-APR-1988; US-179406.
 PR 15-AUG-1988; US-232630.
 PR 23-FEB-1989; US-315342.
 PR 17-OCT-1989; US-422613.
 PR 17-OCT-1989; US-422899.
 PR 22-FEB-1990; US-483913.
 PR 20-AUG-1990; US-569920.
 PR 07-SEP-1990; US-579865.
 PR 18-OCT-1990; US-599543.
 PR 18-OCT-1990; US-600024.
 PA 04-DEC-1990; US-621849.
 PA 04-DEC-1990; US-621988.
 PA 22-FEB-1991; US-660162.
 PA 20-DEC-1991; US-810360.
 PA 28-JAN-1992; US-827052.
 PA 21-FEB-1992; US-841646.
 PI Kuberassampath T, Oppermann H, Ozkaynak E, Pang RHL;
 PI Rueger DC;
 DR WPI: 93-395405/49.
 DR N-PSDB: Q53154.
 PT New pure mammalian osteogenic proteins - induce cartilage and
 PT endochondral bone formation when in association with a matrix
 PS Claim 19: Columns 139-140; 128pp; English.
 CC The osteogenic protein when in association with a matrix can induce
 CC at the locus of an implant the full development cascade of
 CC endochondral bone formation including vascularisation.
 CC mineralisation and bone marrow differentiation. The osteogenic
 CC protein can also be used to repair both bone and cartilage in the
 CC treatment of osteoarthritis. This sequence is the mature form of
 CC the protein.
 SQ Sequence 137 AA;

Query Match 57.7%; Score 56; DB 8; Length 137;
 Best Local Similarity 41.7%; Pred. No. 1.70e+02;
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 56 apqgysayyceg 67
 QY 572 TAQYGHVFCEA 583
 :|:|:|:|:

RESULT 14

ID W95457 standard; protein; 139 AA.
 AC W95457;
 DT 26-MAR-1999 (first entry)
 DE Mature human osteogenic protein 2 (OP2) (residues 264-402).
 KW Cystic kidney disease; renal; therapeutic; osteogenic protein; OP;
 KW bone morphogenic protein; BMP; growth factor-beta superfamily;
 KW polycystic kidney disease; multicystic dysplastic kidney disease;
 KW uraemic medullary cystic disease; human; OP2.
 OS Homo sapiens.
 PN WQ9850061-A1.
 PD 12-NOV-1998.
 PF 06-MAY-1998; U09268.
 PR 07-MAY-1997; US-045909.
 PA (BIOJ) BIOGEN INC.
 PI Gjostrup P, Harris R;
 DR WPI: 99-070084/06.
 PT Treating cystic kidney disease - using renal therapeutic agents or
 PT sequences encoding them, especially from the osteogenic protein/bone
 PT morphogenic protein family
 PS Claim 4: Page -: 67pp; English.
 CC The invention relates to methods for treating cystic kidney diseases. The

CC method comprises administering an effective amount of a renal therapeutic
CC agent or a polynucleotide encoding the therapeutic agent. The therapeutic
CC agent is preferably a soluble or membrane bound polypeptide, e.g. a
CC member of the osteogenic protein/bone morphogenic protein (OP/BMP) family
CC within a transforming growth factor-beta superfamily of proteins. It is
CC especially one of the polypeptides hOP1, hOP1-PP, OPI-16Ser, OPS, OP7,
CC OPI-16Ser, OPI-16Leu, OPI-16Met, OPI-16Ala, OPI-16Val, mOPI, mOPI-PP,
CC hOP2, hOP2-PP, hOP2-Ala, hOP2-Pro, hOP2-Arg, or hOP2-Ser or their
CC biologically active homologues. The method is used to treat humans
CC having, or at risk of, cystic kidney disease, e.g. autosomal recessive
CC (infantile) polycystic disease, multicystic dysplastic kidney disease,
CC uremic medullary cystic disease, and autosomal dominant polycystic
CC kidney disease. The present sequence represents possible mature human
CC osteogenic protein 2 (OP2) that can be used as a therapeutic agent in
CC the method of the invention.
CC Note: This sequence is not provided in the specification. It has been
CC derived from the human OP2 sequence (W95456) obtained from U.S. Patent
CC No. 5,266,683.
SQ Sequence 139 AA;

Query Match 57.7%; Score 56; DB 39; Length 139;
Best Local Similarity 41.7%; Pred. No. 1.70e+02;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 58 apqgysayyceg 69
:::|::|::|
Qy 572 TAQQYGHVFCEA 583

RESULT 15
ID R33400 standard; protein: 139 AA.
AC R33400;
DT 15-JUL-1993 (first entry)
DE Human mature OP-2.
KW morphogenic; osteogenic protein; developmental cascade; human OP-1;
KW hOP-1; mouse OP-1; hOP-2; mOP-2; CBMP2A; CBMP2B;
KW Drosophila DPP; decapentaplegic; Xenopus Vgl; mouse Vgr-1; GDF-1;
KW BMP3; BMP5; BMP6; morphogen 60(A); inflammation; anti-inflammatory;
KW Transforming Growth Factor; TGF-beta super-family.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 43..139
FT /label= C-terminal region
FT /note= "contains conserved 6 Cys skeleton"
PN W09304692-A.
PD 18-MAR-1993.
PF 28-AUG-1992; U07358.
PR 30-AUG-1991; US-752764.
PR 30-AUG-1991; US-752861.
PR 30-AUG-1991; US-753059.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;
PI Pang RHL, Rueger DC, Smart JE;
DR WPI; 93-100652/12.
PT Morphogen-induced modulation of inflammatory response - and
PT resulting tissue damage, e.g. in autoimmune diseases, diabetes,
PT asthma, ischemia reperfusion injury, etc.
PS Claim 26; Page 103-104; 165pp; English.
CC Osteogenic protein (OP)-2 is a preferred morphogen
CC for use in treating tissue damage in e.g. inflammatory disease,
CC autoimmune disease, arthritis, psoriasis, dermatitis, diabetes and
CC emphysema. Proteins having at least 70% homology with OP-1 amino
CC acid sequences can also be used. See R33401 for mOP-2 sequence.
SQ Sequence 139 AA;

Query Match 57.7%; Score 56; DB 7; Length 139;
Best Local Similarity 41.7%; Pred. No. 1.70e+02;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 58 apqgysayyceg 69
:::|::|::|
Qy 572 TAQQYGHVFCEA 583

Search completed: Thu Jul 8 19:07:56 1999
Job time : 18 secs.

WIRE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:06:31 1999; MasPar time 3.53 Seconds

Tabular output not generated. 136.265 Million cell updates/sec

Title: >US-09-041-236-2
Description: (572-583) from US09041236.pap (29 of 45)
Perfect Score: 97
Sequence: 1 TAAQYGHVFCEA 12

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 25.321; Variance 43.270; scale 0.585

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Length	ID	Description	Pred. No.
1	62	63.9	174	2 C64053 flavodoxin A - Haemop	2.25e+00
2	62	63.9	176	2 A37319 flavodoxin A - Escher	2.25e+00
3	60	61.9	97	2 JMW013 plastocyanin - green	4.69e+00
4	60	61.9	174	2 A70828 hypothetical protein	4.69e+00
5	57	58.8	157	2 S58073 Probable olfactory re	1.38e+01
6	57	58.8	238	2 S44793 F09G8.2 protein - Cae	1.38e+01
7	57	58.8	464	2 S12591 beta-1-adrenergic rec	1.38e+01
8	57	58.8	466	2 S36794 beta-1-adrenergic rec	1.38e+01
9	57	58.8	477	1 QRHUB1 beta-1-adrenergic rec	1.38e+01
10	57	58.8	480	2 I53053 beta 1 adrenergic rec	1.38e+01
11	56	57.7	79	2 A28840 Ig kappa chain V regi	1.96e+01
12	56	57.7	101	2 S20810 Ig kappa chain V regi	1.96e+01
13	56	57.7	377	2 S77142 tryptophan--trna liga	1.96e+01
14	56	57.7	402	2 A45056 osteogenic protein 2	1.96e+01
15	55	56.7	98	1 CUKLCF plastocyanin - Chlore	2.77e+01
16	55	56.7	157	2 S58037 Probable olfactory re	2.77e+01
17	55	56.7	292	2 S25818 hypothetical protein	2.77e+01
18	55	56.7	309	2 S1336 Olfactory receptor -	2.77e+01
19	55	56.7	333	2 A31923 amalgam protein precu	2.77e+01
20	55	56.7	342	2 E71359 hypothetical protein	2.77e+01
21	55	56.7	404	2 A47425 serotonin receptor 7	2.77e+01
22	55	56.7	445	2 A48881 serotonin receptor 7	2.77e+01
23	55	56.7	448	2 S40687 serotonin receptor 7	2.77e+01

24	55	56.7	448	2 A47519 serotonin receptor 7	2.77e+01
25	55	56.7	448	2 S36402 serotonin receptor 7	2.77e+01
26	55	56.7	487	2 S41951 UTP-glucose glucosylt	2.77e+01
27	55	56.7	831	2 S44843 K06H7.3 protein - Cae	2.77e+01
28	55	56.7	1203	2 A33165 finger protein sdc-1	2.77e+01
29	54	55.7	69	2 PH1080 Ig light chain V regi	3.91e+01
30	54	55.7	81	2 S22024 Ig kappa chain - mous	3.91e+01
31	54	55.7	102	2 S26346 Ig kappa chain V regi	3.91e+01
32	54	55.7	104	2 B43413 Ig kappa chain V regi	3.91e+01
33	54	55.7	106	2 PL0267 Ig kappa chain V regi	3.91e+01
34	54	55.7	107	2 B45722 anti-glycoprotein H m	3.91e+01
35	54	55.7	107	2 A45722 anti-glycoprotein H m	3.91e+01
36	54	55.7	123	2 S35479 Ig kappa chain precu	3.91e+01
37	54	55.7	138	2 A26471 Ig kappa chain precu	3.91e+01
38	54	55.7	147	2 A40266 cardiac calcium chann	3.91e+01
39	54	55.7	197	2 A44363 voltage-gated dihydro	3.91e+01
40	54	55.7	342	2 S12955 calcium channel prote	3.91e+01
41	54	55.7	622	2 S57684 probable membrane pro	3.91e+01
42	54	55.7	668	1 Q0BEW1 UL52 protein - human	3.91e+01
43	54	55.7	1144	2 A36968 p1-like adhesin precu	3.91e+01
44	54	55.7	1864	1 JQ1657 genome polyprotein B	3.91e+01
45	54	55.7	2143	2 JH0427 voltage-dependent cal	3.91e+01

ALIGNMENTS

RESULT 1
ENTRY #type complete
TITLE flavodoxin A - Haemophilus influenzae (strain Rd KW20)
ORGANISM #formal_name Haemophilus influenzae
DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 10-Oct-1997
ACCESSIONS C64053
REFERENCE A64000
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal Science (1995) 269:496-512
#title Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
#cross-references MUID:95350630
#accession C64053
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-174 #label TIGR
#cross-references TIGR:U32704; GB:L42023; NID:gi573143; PID:gi573149; TIGR:HI0191
#note named as homolog to a protein from Escherichia coli
CLASSIFICATION #superfamily flavodoxin; flavodoxin homology
KEYWORDS electron transfer; flavoprotein; FMN
FEATURE 6-165
SUMMARY #domain flavodoxin homology #label FLX
#length 174 #molecular-weight 19627 #checksum 8175
Query Match 63.9%; Score 62; DB 2; Length 174;
Best Local similarity 50.0%; Pred. No. 2.25e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 92 EDYADVFCEA 101

QY 574 QQYGHVFCEA 583

RESULT 2

ENTRY #type complete
TITLE flavodoxin A - Escherichia coli

```

ALTERNATE_NAMES  flavodoxin 1
ORGANISM          #formal_name Escherichia coli
DATE             11-Sep-1992 #sequence_revision 11-Sep-1992 #text_change
17-Jul-1998
ACCESSIONS       A37319; C64803
REFERENCE        A37319
#authors         Osborne, C.; Chen, L.M.; Matthews, R.G.
#journal         J. Bacteriol. (1991) 173:1729-1737
#title           Isolation, cloning, mapping, and nucleotide sequencing of the
#                gene encoding flavodoxin in Escherichia coli.
#cross-references MUID:91154129
#accession       A37319
##status         preliminary
##molecule_type DNA
##residues       1-176 #label OSB
##cross-references GB:M59426; NID:g145985; PID:g145986
REFERENCE        A64720
#authors         Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                Y.
#journal         Science (1997) 277:1453-1462
#title           The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession       C64803
##status         nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues       1-176 #label BLAT
##cross-references GB:AE000172; GB:U000096; NID:g1786896; PID:g1786900;
                UWGP:b0684
##experimental_source strain K-12, substrain MG1655
GENETICS
#gene            fldA
#map_position    15.9 min
CLASSIFICATION   #superfamily flavodoxin; flavodoxin homology
KEYWORDS         electron transfer; flavoprotein; FMN
FEATURE          6-165
#domain          flavodoxin homology #label FLX
SUMMARY          #length 176 #molecular-weight 17377 #checksum 9726

Query Match      63.9%; Score 62; DB 2; Length 176;
Best Local Similarity 50.0%; Pred. No. 2.25e+00;
Matches          5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db              92 EDYAEVFCDA 101
QY              574 QQYGHVFCEA 583

RESULT          3
ENTRY           JW0013 #type complete
TITLE           plastocyanin - green alga (Scenedesmus obliquus)
ORGANISM        #formal_name Scenedesmus obliquus
DATE            30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
23-Feb-1997
ACCESSIONS      JW0013; A38820
REFERENCE        JW0013
#authors         Kelly, J.M.; Ambler, R.P.
#citation         unpublished results, cited by Sykes, A.G., Chem. Soc. Rev.
                14, 283-315, 1985
#description     Structure and electron-transfer reactivity of the blue copper
                protein plastocyanin.
#accession       JW0013
##molecule_type protein
##residues       1-92, 'T', 95-97 #label KEL
REFERENCE        A38820
#authors         Ambler, R.P.
#citation         unpublished results, cited by McGinnis, J., Sinclair-Day,
                J.D., Sykes, A.G., Poulos, R., Moore, J., and Wright, P.D.,
                Inorg. Chem. 27, 2306-2312, 1988
#accession       A38820
##molecule_type protein

```

```

##residues       92-97 #label AMB
CLASSIFICATION   #superfamily plastocyanin
KEYWORDS         chloroplast; electron transfer
FEATURE          37,82,85,90
#binding_site    copper (His, Cys, His, Met) (type 1)
SUMMARY          #length 97 #molecular-weight 10105 #checksum 5566
                #status experimental
Query Match      61.9%; Score 60; DB 2; Length 97;
Best Local Similarity 66.7%; Pred. No. 4.69e+00;
Matches          8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db              74 TAGEYG-YFCEP 84
QY              572 TAAQYGHVFCEA 583

RESULT          4
ENTRY           A70828 #type complete
TITLE           hypothetical protein Rv0461 - Mycobacterium tuberculosis
                (strain H37RV)
ORGANISM        #formal_name Mycobacterium tuberculosis
DATE            17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
A70828
REFERENCE        A70500
#authors         Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                III, C.E.; rekaia, F.; Badcock, K.; Busham, D.; Brown, D.;
                Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
                Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal         Nature (1998) 393:537-544
#title           Deciphering the biology of Mycobacterium tuberculosis from
                the complete genome sequence.
#cross-references MUID:98295987
#accession       A70828
##status         preliminary; nucleic acid sequence not shown;
                translation not shown
##molecule_type DNA
##residues       1-174 #label COL
##cross-references GB:AL021933; GB:AL123456; NID:g3261529; PID:e1252531;
                PID:g2909537
##experimental_source strain H37Rv
GENETICS
#gene            Rv0461
SUMMARY          #length 174 #molecular-weight 18622 #checksum 4031

Query Match      61.9%; Score 60; DB 2; Length 174;
Best Local Similarity 54.5%; Pred. No. 4.69e+00;
Matches          6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db              152 AAPLYGYVYCE 162
QY              572 TAAQYGHVFCE 582

RESULT          5
ENTRY           S58073 #type fragment
TITLE           probable olfactory receptor tpor12 - rat (fragment)
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            13-Jan-1996 #sequence_revision 13-Mar-1997 #text_change
10-Sep-1997
ACCESSIONS      S58073
REFERENCE        S57995
#authors         Vanderhaeghen, P.; Schurmans, S.; Vassart, G.; Parmentier, M.
#citation         submitted to the EMBL Data Library, July 1995
                Male germ cells from several mammalian species express a
                specific repertoire of olfactory receptor genes.

```

```
#accession S58073 preliminary
##status preliminary
##molecule_type mRNA
##residues 1-157 ##label VAN
##cross-references EMBL:X89700; NID:g902716; PID:g902717
SUMMARY #length 157 #checksum 6057

Query Match 58.8%; Score 57; DB 2; Length 157;
Best Local Similarity 45.5%; Pred. No. 1.38e+01;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 45 SSOEIDHFFCE 55
QY 572 TAOQYGHYFCE 582

RESULT 6
ENTRY S44793 #type complete
TITLE F09G8.2 protein - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 14-Sep-1994 #sequence_revision 12-May-1995 #text_change
09-Sep-1997
ACCESSIONS S44793
REFERENCE S44774
#authors Anderson, K.
#submission submitted to the EMBL Data Library, February 1993
#description Sequence of the C. elegans cosmid F09G8.
#accession S44793
##status preliminary
##molecule_type DNA
##residues 1-238 ##label AND
##cross-references EMBL:L11247; NID:gl56280; PID:gl56283
GENETICS
#introns 32/2; 171/3
SUMMARY #length 238 #molecular-weight 26992 #checksum 1042

Query Match 58.8%; Score 57; DB 2; Length 238;
Best Local Similarity 46.2%; Pred. No. 1.38e+01;
Matches 6; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

Db 142 TAEKYGOSFFCAS 154
QY 572 TAOQYGH-YFCEA 583

RESULT 7
ENTRY S12591 #type complete
TITLE beta-1-adrenergic receptor - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
08-Sep-1997
ACCESSIONS S12591; S40185; A36618
REFERENCE S12591
#authors Shimomura, H.; Terada, A.
#journal Nucleic Acids Res. (1990) 18:4591
#title Primary structure of the rat beta-1 adrenergic receptor gene.
#cross-references EMBL:03056399
#accession S12591
##molecule_type DNA
##residues 1-464 ##label SHI
##cross-references EMBL:D00634
REFERENCE S40185
#authors Shimomura, H.; Terada, A.
#submission submitted to the EMBL Data Library, July 1991
#accession S40185
##molecule_type DNA
##residues 1-25; 'A', '26-174', 'R', '175-464' ##label SH2
##cross-references EMBL:D00634; NID:g220670; PID:d1000982; PID:g220671
REFERENCE A36618
#authors Machida, C.A.; Bunzow, J.R.; Searles, R.P.; Van Tol, H.;
Tester, B.; Neve, K.A.; Teal, P.; Nipper, V.; Civelli, O.
#journal J. Biol. Chem. (1990) 265:12960-12965
#title Molecular cloning and expression of the rat beta-1-adrenergic
receptor gene.
#accession A36618
#submission submitted to the EMBL Data Library, July 1991
#accession A36618
##molecule_type mRNA
##residues 1-477 ##label FRI
##cross-references GB:J03019; NID:gl178199; PID:gl178200
GENETICS
#gene GDB:ADRB1; ADRB1R
```

```
receptor gene.
#cross-references MUID:90330633
#accession A36618
##molecule_type DNA
##residues 1-26; 'A', '27-160', 'L', '162-174', 'R', '175-264', 'T', '266-464'
##label MAC
##cross-references GB:J05561; NID:g203073; PID:g203074
CLASSIFICATION #superfamily vertebrate rhodopsin
KEYWORDS G protein-coupled receptor; glycoprotein; phosphoprotein;
transmembrane protein
SUMMARY #length 464 #molecular-weight 50204 #checksum 2598

Query Match 58.8%; Score 57; DB 2; Length 464;
Best Local Similarity 62.5%; Pred. No. 1.38e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 124 EYGSFFCE 131
QY 575 QYGHYFCE 582

RESULT 8
ENTRY S36794 #type complete
TITLE beta-1-adrenergic receptor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 09-Dec-1993 #sequence_revision 10-Nov-1995 #text_change
17-Mar-1999
ACCESSIONS S36794
REFERENCE S36794
#authors Jasper, J.R.; Link, R.E.; Chruscinski, A.J.; Kobilka, B.K.;
Bernstein, D.
#journal Biochim. Biophys. Acta (1993) 1178:307-309
#title Primary structure of the mouse beta(1)-adrenergic receptor
gene.
#cross-references MUID:93372116
#accession S36794
##status preliminary
##molecule_type DNA
##residues 1-466 ##label JAS
##cross-references EMBL:L10084; NID:g293278; PID:g293279
CLASSIFICATION #superfamily vertebrate rhodopsin
KEYWORDS G protein-coupled receptor; phosphoprotein; transmembrane
protein
SUMMARY #length 466 #molecular-weight 50479 #checksum 2043

Query Match 58.8%; Score 57; DB 2; Length 466;
Best Local Similarity 62.5%; Pred. No. 1.38e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 125 EYGSFFCE 132
QY 575 QYGHYFCE 582

RESULT 9
ENTRY QRHUB1 #type complete
TITLE beta-1-adrenergic receptor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 22-Jan-1993 #sequence_revision 18-Aug-1995 #text_change
18-Sep-1998
ACCESSIONS A39911
REFERENCE A39911
#authors Friele, T.; Collins, S.; Daniel, K.W.; Caron, M.G.;
Lefkowitz, R.J.; Kobilka, B.K.
#journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:7920-7924
#title Cloning of the cDNA for the human beta-1-adrenergic receptor.
#cross-references MUID:88068509
#accession A39911
##molecule_type mRNA
##residues 1-477 ##label FRI
##cross-references GB:J03019; NID:gl178199; PID:gl178200
GENETICS
#gene GDB:ADRB1; ADRB1R
```

```
#cross-references GDB:119654; OMIM:109630
map_position 10q25-10q25
CLASSIFICATION #superfamily vertebrate rhodopsin
KEYWORDS G protein-coupled receptor; glycoprotein; transmembrane protein
FEATURE
58-84 #domain transmembrane #status predicted #label TM1\
94-121 #domain transmembrane #status predicted #label TM2\
133-154 #domain transmembrane #status predicted #label TM3\
177-199 #domain transmembrane #status predicted #label TM4\
224-245 #domain transmembrane #status predicted #label TM5\
325-346 #domain transmembrane #status predicted #label TM6\
358-377 #domain transmembrane #status predicted #label TM7\
15 #binding_site carbohydrate (Asn) (covalent) #status predicted
SUMMARY #length 477 #molecular-weight 51223 #checksum 1819
Query Match 58.8%; Score 57; DB 1; Length 477;
Best Local Similarity 62.5%; Pred. No. 1.38e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 125 EYGSFCE 132
QY 575 QYGHYFCE 582

RESULT 10
ENTRY #type complete
TITLE beta 1 adrenergic receptor - rhesus macaque
ORGANISM #formal_name Macaca mulatta #common_name rhesus macaque
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 30-May-1997
ACCESSIONS I53053
REFERENCE I53053
#authors Searles, R.P.; Nipper, V.J.; Machida, C.A.
#journal DNA Seq. (1994) 4:231-241
#title The rhesus macaque beta 1-adrenergic receptor gene: structure of the gene and comparison of the flanking sequences with the rat beta 1-adrenergic receptor gene.
#cross-references MUID:95078456
#accession I53053
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 1-480 #label RES
#cross-references EMBL:X75540; NID:G510532; PID:g510533
CLASSIFICATION #superfamily vertebrate rhodopsin
KEYWORDS neurotransmitter receptor
SUMMARY #length 480 #molecular-weight 51608 #checksum 2585
Query Match 58.8%; Score 57; DB 2; Length 480;
Best Local Similarity 62.5%; Pred. No. 1.38e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 125 EYGSFCE 132
QY 575 QYGHYFCE 582

RESULT 11
ENTRY #type fragment
TITLE Ig kappa chain V region (HP21) - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change 20-Mar-1998
ACCESSIONS A28840; M25114
REFERENCE A91028
#authors Ollier, P.; Rocca-Serra, J.; Somme, G.; Theze, J.; Fougereau, M.
#journal EMBO J. (1985) 4:3681-3688
#title The idiotypic network and the internal image: possible regulation of a germ-line network by paucigene encoded Ab2 (anti-idiotypic) antibodies in the GAT system.
#cross-references MUID:86136012

#accession A28840
#molecule_type mRNA
#residues 1-79 #label OLL
#cross-references GB:X03387; NID:g52168; PID:g52169
CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology
KEYWORDS heterotrimer; immunoglobulin
SUMMARY #length 79 #checksum 5088
Query Match 57.7%; Score 56; DB 2; Length 79;
Best Local Similarity 40.0%; Pred. No. 1.96e+01;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 52 PEDFGSYFC 61
QY 573 AQYGHYFCE 582

RESULT 12
ENTRY #type complete
TITLE Ig kappa chain V region (hybridoma C8) - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Sep-1997
ACCESSIONS S20810
REFERENCE S20809
#authors Hooenboom, H.; Dubois, P.; Raus, J.; Volckaert, G.
#submission submitted to the EMBL Data Library, September 1990
#description Nucleotide sequences of the variable region cDNAs encoding a murine antibody with specificity for the human IL-2 receptor.
#accession S20810
#status preliminary
#molecule_type mRNA
#residues 1-101 #label HOO
#cross-references EMBL:X54693; NID:g50251; PID:g50252
CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology
KEYWORDS heterotrimer; immunoglobulin
SUMMARY #length 101 #molecular-weight 11082 #checksum 5120
Query Match 57.7%; Score 56; DB 2; Length 101;
Best Local Similarity 30.0%; Pred. No. 1.96e+01;
Matches 3; Conservative 7; Mismatches 0; Indels 0; Gaps 0;
Db 73 PEDFGRYCQ 82
QY 573 AQYGHYFCE 582

RESULT 13
ENTRY #type complete
TITLE tryptophan--tRNA ligase (EC 6.1.1.2) - Synechocystis sp. (strain PCC 6803)
ALTERNATE_NAMES protein slr1884; tryptophanyl-tRNA synthetase
ORGANISM #formal_name Synechocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
ACCESSIONS S77142
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
#cross-references MUID:97061201
#accession S77142
#status nucleic acid sequence not shown; translation not shown
```


##molecule_type DNA
##residues 1-377 ##label KAN
##cross-references EMBL:D80908; GB:AB001339; NID:g1652725; PID:d1018433;
PID:g1652781
##note the nucleotide sequence was submitted to the EMBL Data Library, June 1996

GENETICS

##gene trps
CLASSIFICATION #superfamily tryptophan--trna ligase
KEYWORDS aminoacyl-trna synthetase; ligase; protein biosynthesis
SUMMARY #length 377 #molecular-weight 42367 #checksum 6919

Query Match 57.7%; Score 56; DB 2; Length 377;
Best Local Similarity 50.0%; Pred. No. 1.96e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 73 QHYDNFFC 80
I: I : I I I
QY 574 QQYGHYFC 581

RESULT 14

ENTRY A45056 #type complete
TITLE osteogenic protein 2 precursor - human
ALTERNATE_NAMES bone morphogenetic protein 8; OP-2
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
09-Apr-1998
ACCESSIONS A45056
REFERENCE A45056
#authors Oozkaynak, E.; Schneegelsberg, P.N.; Jin, D.F.; Clifford, G.M.; Warren, F.D.; Drier, E.A.; Oppermann, H.
#journal J. Biol. Chem. (1992) 267:25220-25227
#title Osteogenic protein-2. A new member of the transforming growth factor-beta superfamily expressed early in embryogenesis.
#cross-references MUID:93094231

#accession A45056
##status preliminary; not compared with conceptual translation
##molecule_type nucleic acid
##residues 1-402 ##label OZK
##cross-references GB:M97016; NID:g189389; PID:g189390
##experimental_source hippocampus
##note sequence extracted from NCBI backbone (NCBIP:120189)

GENETICS

##gene GDB:BMP8; OP-2
##cross-references GDB:136392
CLASSIFICATION #superfamily inhibin
SUMMARY #length 402 #molecular-weight 44764 #checksum 5993

Query Match 57.7%; Score 56; DB 2; Length 402;
Best Local Similarity 41.7%; Pred. No. 1.96e+01;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 321 APOGYSAYYCEG 332
I: I : I I I
QY 572 TAAQYGHYFCEA 583

RESULT 15

ENTRY CUKLCF #type complete
TITLE plastocyanin - Chlorella fusca
ORGANISM #formal_name Chlorella fusca
DATE 07-May-1981 #sequence_revision 07-May-1981 #text_change
20-Feb-1998
ACCESSIONS A00310
REFERENCE A00310
#authors Kelly, J.; Ambler, R.P.
#journal Biochem. J. (1974) 143:681-690
#title The amino acid sequence of plastocyanin from Chlorella fusca.
#cross-references MUID:75183336
#accession A00310
##molecule_type protein
##residues 1-98 ##label KEL

CLASSIFICATION #superfamily plastocyanin
KEYWORDS chloroplast; copper; electron transfer; metalloprotein
FEATURE 38,83,86,91 #binding_site copper (His, Cys, His, Met) (type 1)
#status predicted
SUMMARY #length 98 #molecular-weight 10299 #checksum 9528

Query Match 56.7%; Score 55; DB 1; Length 98;
Best Local Similarity 66.7%; Pred. No. 2.77e+01;
Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Db 75 TAGTYG-YFCEP 85
I: I : I I I I I
QY 572 TAAQYGHYFCEA 583

Search completed: Thu Jul 8 19:06:43 1999
Job time : 12 secs.

THIS PAGE BLANK (USPTO)

WQSRH (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:06:00 1999; Maspar time 2.38 Seconds
Tabular output not generated. 142.674 Million cell updates/sec

Title: >US-09-041-236-2
Description: (572-583) from US09041236.pep (29 of 45)
Perfect Score: 97
Sequence: 1 TAQYGHYFCEA 12

Scoring table:
PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 25.767; Variance 37.257; scale 0.692

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	62	63.9	173	1	FLAV_HAEIN	4.01e-01
2	62	63.9	175	1	FLAV_ECOLI	4.01e-01
3	60	61.9	97	1	PLAS_SCEOB	9.40e-01
4	57	58.8	238	1	YLS2_CAEEL	3.26e+00
5	57	58.8	346	1	B1AR_SHEEP	3.26e+00
6	57	58.8	466	1	B1AR_RAT	3.26e+00
7	57	58.8	466	1	B1AR_MOUSE	3.26e+00
8	57	58.8	466	1	B1AR_PIG	3.26e+00
9	57	58.8	477	1	B1AR_HUMAN	3.26e+00
10	57	58.8	480	1	B1AR_MACMU	3.26e+00
11	56	57.7	377	1	SYW_SYNY3	4.89e+00
12	56	57.7	399	1	BMA_MOUSE	4.89e+00
13	56	57.7	402	1	BMP8_HUMAN	4.89e+00
14	56	57.7	879	1	FPRP_RAT	4.89e+00
15	55	56.7	98	1	PLAS_CHLFU	7.30e+00
16	55	56.7	292	1	YKVA_YEAST	7.30e+00
17	55	56.7	333	1	AMAL_DROME	7.30e+00
18	55	56.7	445	1	5-HYDROXYTRYPTAMINE 7	7.30e+00
19	55	56.7	446	1	5HT_CAVPO	7.30e+00
20	55	56.7	448	1	5HT_RAT	7.30e+00
21	55	56.7	448	1	5HT_MOUSE	7.30e+00
22	55	56.7	487	1	UF05_MANES	7.30e+00
23	55	56.7	831	1	YMK3_CAEEL	7.30e+00

24	55	56.7	1203	1	SDC1_CAEEL	7.30e+00
25	54	55.7	338	1	LAMP_HUMAN	1.09e+01
26	54	55.7	338	1	LAMP_CHICK	1.09e+01
27	54	55.7	338	1	LAMP_RAT	1.09e+01
28	54	55.7	484	1	UL32_HSV6U	1.09e+01
29	54	55.7	485	1	UL32_HSV7J	1.09e+01
30	54	55.7	526	1	UBPW_MOUSE	1.09e+01
31	54	55.7	622	1	YGA4_YEAST	1.09e+01
32	54	55.7	668	1	UL52_HCMVA	1.09e+01
33	54	55.7	1864	1	VGNB_RCMV	1.09e+01
34	54	55.7	2169	1	CICC_RAT	1.09e+01
35	53	54.6	145	1	PLAS_CHLRE	1.60e+01
36	53	54.6	259	1	YL23_YEAST	1.60e+01
37	53	54.6	331	1	GPRI_HUMAN	1.60e+01
38	53	54.6	343	1	Z183_HUMAN	1.60e+01
39	53	54.6	739	1	PEC1_BOVIN	1.60e+01
40	53	54.6	1259	1	CAML_RAT	1.60e+01
41	53	54.6	1260	1	CAML_MOUSE	1.60e+01
42	52	53.6	353	1	CEPU_CHICK	2.36e+01
43	52	53.6	410	1	ARPI_CHICK	2.36e+01
44	52	53.6	917	1	AKH2_MAIZE	2.36e+01
45	52	53.6	1648	1	YJ9H_YEAST	2.36e+01

ALIGNMENTS

RESULT	1					
ID	FLAV_HAEIN	STANDARD:	PRT:	173 AA.		
AC	P44562:					
DT	01-NOV-1995 (REL. 32, CREATED)					
DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)					
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)					
DE	FLAVODOXIN.					
GN	FLDA OR HI0191.					
OS	HAEMOPHILUS INFLUENZAE.					
OC	BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;					
OC	HAEMOPHILUS.					
RN	[1]					
RC	SEQUENCE FROM N.A.					
RP	STRAIN-RD / KW20:					
RX	MEDLINE: 95350630.					
RA	FLITSCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,					
RA	KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,					
RA	MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,					
RA	SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,					
RA	WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,					
RA	UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,					
RA	FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEORHAGEN N.S.M.,					
RA	GNEM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,					
RA	VENTER J.C.;					
RT	"Whole-genome random sequencing and assembly of Haemophilus					
RT	influenzae Rd.,"					
RL	SCIENCE 269:496-512(1995).					
CC	-!- FUNCTION: LOW-POTENTIAL ELECTRON DONOR TO A NUMBER OF REDOX					
CC	ENZYMES (BY SIMILARITY).					
CC	-!- COFACTOR: FMN (BY SIMILARITY).					
CC	-!- SIMILARITY: BELONGS TO THE FLAVODOXIN FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; U32704; G1573149; -					
DR	TIGR; HI0191; -					
DR	PROSITE; PS00201; FLAVODOXIN; 1.					
DR	PFAM; PF00258; flavodoxin; 1.					
DR	HSP; P23243; LAHN.					
KW	ELECTRON TRANSPORT; FLAVOPROTEIN; FMN.					
FT	INIT_MET 0 0 BY SIMILARITY.					

SQ SEQUENCE 173 AA; 19495 MW; EEE97CFD CRC32;
Query Match 63.9%; Score 62; DB 1; Length 173;
Best Local Similarity 50.0%; Pred. No. 4.01e-01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 91 EDYADYFCDA 100
Qy 574 QOYGHYFCEA 583

RESULT 2
ID FLAV_ECOLI STANDARD; PRT; 175 AA.
AC P23243;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE FLAVODOXIN 1.
GN FLDA
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.
RX MEDLINE; 91154129.
RA OSBORNE C., CHEN L.-M., MATTHEWS R.G.;
RT "Isolation, cloning, mapping, and nucleotide sequencing of the gene
encoding flavodoxin in *Escherichia coli*.";
RL J. BACTERIOL. 173:1729-1737(1991).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL SCIENCE 277:1453-1474(1997).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN-K12;
RX MEDLINE; 97061202.
RA OSHIMA T., AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,
RA IKEMOTO K., INADA T., ITOH T., KAJIHARA M., KANAI K., KASHIMOTO K.,
RA KIMURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K.,
RA MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SAITO N.,
RA SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,
RA YANO M., HORIUCHI T.;
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA RES. 3:137-155(1996).
RN [4]
RN SEQUENCE OF 1-10.
RP MEDLINE; 95050480.
RA JENKINS C.M., WATERMAN M.R.;
RT "Flavodoxin and NADPH-flavodoxin reductase from *Escherichia coli*
support bovine cytochrome P450c17 hydroxylase activities.";
RL J. BIOL. CHEM. 269:27401-27408(1994).
RN [5]
RN STRUCTURE BY NMR.
RP MEDLINE; 97234567.
RA PONTING H., OTTING G.;
RT "NMR assignments, secondary structure and hydration of oxidized
Escherichia coli flavodoxin.";
RL EUR. J. BIOCHEM. 244:384-399(1997).
RN [6]
RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE; 98078562.
RA HOOVER D.M., LUDWIG M.L.;
RT "A flavodoxin that is required for enzyme activation: the structure
of oxidized flavodoxin from *Escherichia coli* at 1.8-A resolution.";
RL PROTEIN SCI. 6:2525-2537(1997).

CC -!- FUNCTION: LOW-POTENTIAL ELECTRON DONOR TO A NUMBER OF REDOX
CC ENZYMES (POTENTIAL).
CC -!- COFACTOR: FMN.
CC -!- SIMILARITY: BELONGS TO THE FLAVODOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M59426; G145986; -.
DR EMBL; AE000172; G1786900; -.
DR EMBL; D90707; G1651291; -.
DR EMBL; D90708; G1651296; -.
DR FIR; A37319; A37319.
DR PDB; LAHN; 10-DEC-97.
DR PDB; LAG9; 24-DEC-97.
DR ECO2DBASE; A019.0; 6TH EDITION.
DR ECOGENE; EG10318; FLDA.
DR PROSITE; PS00201; FLAVODOXIN; 1.
DR PFAM; PF00258; flavodoxin; 1.
KW ELECTRON TRANSPORT; FLAVOPROTEIN; FMN; 3D-STRUCTURE.
FT INIT_MET 0
SQ SEQUENCE 175 AA; 19606 MW; CC6F305F CRC32;
Query Match 63.9%; Score 62; DB 1; Length 175;
Best Local Similarity 50.0%; Pred. No. 4.01e-01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 91 EDYADYFCDA 100
Qy 574 QOYGHYFCEA 583

RESULT 3
ID PIAS_SCOB STANDARD; PRT; 97 AA.
AC P26956;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PLASTOCYANIN.
OS SCENEDESMUS OBLIQUUS.
OC EUKARYOTA; VIRIDIPANTAE; CHLOROPHYTA; CHLOROPHYCEAE; CHLOROCOCCALES;
OC SCENEDESMACEAE; SCENEDESMUS.
[1]
RN SEQUENCE.
RA KELLY J.M., AMBLER R.P.;
RT "Structure and electron-transfer reactivity of the blue copper protein
plastocyanin.";
RL UNPUBLISHED RESULTS, CITED BY:
RL SYKES A.G.;
RL CHEM. SOC. REV. 14:283-315(1985).
[2]
RN REVISIONS TO C-TERMINAL.
RA AMBLER R.P.;
RT UNPUBLISHED RESULTS, CITED BY:
RL MCGINNIS, J., SINCLAIR-DAY, J.D., SYKES A.G., POWLS, R., MOORE, J.,
RL WRIGHT P.D.;
RL INORG. CHEM. 27:2306-2312(1988).
[3]
RN STRUCTURE BY NMR.
RX MEDLINE; 88178108.
RA MOORE J.M., CASE D.A., CHAZIN W.J., GIPPERT G.P., HAVEL T.F.,
RA POWLS R., WRIGHT P.E.;
RT "Three-dimensional solution structure of plastocyanin from the green
alga *Scenedesmus obliquus*.";
RL SCIENCE 240:314-317(1988).
[4]
RN STRUCTURE BY NMR.
RX MEDLINE; 89088156.

RA MOORE J.M., CHAZIN W.J., POWLS R., WRIGHT P.E.;
RT "1H NMR studies of plastocyanin from Scenedesmus obliquus: complete
sequence-specific assignment, secondary structure analysis, and
global fold."
RL BIOCHEMISTRY 27:7806-7816(1988).
CC -!- FUNCTION: PLASTOCYANIN PARTICIPATES IN ELECTRON TRANSFER BETWEEN
CC P700 AND THE CYTOCHROME B/F COMPLEX IN PHOTOSYSTEM I.
CC -!- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
CC MEMBRANE SURFACE IN CHLOROPLASTS.
CC -!- SIMILARITY: CONTAINS ONE PLASTOCYANIN-LIKE DOMAIN.
DR PIR: JWO013; JWO013.
DR PROSITE: PS00196; COPPER_BLUE; 1.
DR PFAM: PF00127; copper_bind; 1.
DR HSSP: P18068; 2PLT.
KW CHLOROPLAST; ELECTRON TRANSPORT; COPPER; THYLAKOID MEMBRANE.
FT DOMAIN 1 97 PLASTOCYANIN-LIKE.
FT METAL 37 37 COPPER (BY SIMILARITY).
FT METAL 82 82 COPPER (BY SIMILARITY).
FT METAL 85 85 COPPER (BY SIMILARITY).
FT METAL 90 90 COPPER (BY SIMILARITY).
SQ SEQUENCE 97 AA; 10105 MW; 88A421BE CRC32;

Query Match 61.9%; Score 60; DB 1; Length 97;
Best Local Similarity 66.7%; Pred. No. 9.40e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 74 TAGGYG-YFCEP 84
||: ||: ||: ||:
QY 572 TAQGYGHYFCEA 583

RESULT 4
ID YLS2_CABEL STANDARD; PRT: 238 AA.
AC P34387;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 27.0 KD PROTEIN F09G8.2 IN CHROMOSOME III.
GN F09G8.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEN R.,
RA SIMS M., SHALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WORLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL NATURE 368:32-38(1994).
CC -!- SIMILARITY: BELONGS TO THE DNASE II FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L11247; G156283; -
DR PIR: S44793; S44793.
DR WORMPEP: F09G8.2; CE00137.

KW HYPOTHETICAL PROTEIN; HYDROLASE.
SQ SEQUENCE 238 AA; 26992 MW; 4FCBC49F CRC32;

Query Match 58.8%; Score 57; DB 1; Length 238;
Best Local Similarity 46.2%; Pred. No. 3.26e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

Db 142 TABKYQGSFFCAS 154
||: ||: ||: ||:
QY 572 TAQGYGH-YFCEA 583

RESULT 5
ID BLAR_SHEEP STANDARD; PRT: 346 AA.
AC Q28927;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE BETA-1 ADRENERGIC RECEPTOR (FRAGMENT).
GN ADRB1.
OS OVIS ARIES (SHEEP).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; CAPRINAE; OVIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96187488.
RA PADBURY J.F., TSENG Y.T., WASCHER J.A.;
RT "A cloning strategy for G-protein-coupled hormone receptors: the
RT ovine beta 1-adrenergic receptor."
RL REPROD. FERTIL. DEV. 7:521-525(1995).
CC -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
CC INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
CC PROTEINS. THIS RECEPTOR BINDS EPINEPHRINE AND NOREPINEPHRINE WITH
CC APPROXIMATIVELY EQUAL AFFINITY.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- PTM: HOMOLOGOUS DESENSITIZATION OF THE RECEPTOR IS MEDIATED BY ITS
CC PHOSPHORYLATION BY BETA-ADRENERGIC RECEPTOR KINASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S81783; G1438750; -
DR GCRDB; GCR_1901; -
DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
DR PFAM: PF00001; 7tm_1; 1.
DR HSSP: P07700; 1DSP.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT NON_TER 1 1
FT DOMAIN <1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 31 2 (POTENTIAL).
FT DOMAIN 32 42 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 43 66 3 (POTENTIAL).
FT DOMAIN 67 86 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 87 110 4 (POTENTIAL).
FT DOMAIN 111 132 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 133 156 5 (POTENTIAL).
FT DOMAIN 157 223 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 224 247 6 (POTENTIAL).
FT DOMAIN 248 254 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 255 278 7 (POTENTIAL).
FT DOMAIN 279 346 CYTOPLASMIC (POTENTIAL).
FT DISULFID 42 120 BY SIMILARITY.
FT LIPID 290 290 PALMITATE (BY SIMILARITY).
SQ SEQUENCE 346 AA; 38157 MW; 05FA9B41 CRC32;

```
Query Match      58.8%; Score 57; DB 1; Length 346;
Best Local Similarity 62.5%; Pred. No. 3.26e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 36 EYGSFFCE 43
QY 575 QYGHYFCE 582

RESULT 6
ID BLAR_RAT STANDARD; PRT; 466 AA.
AC P18090;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DE BETA-1 ADRENERGIC RECEPTOR.
DE BETA-1 ADRENERGIC RECEPTOR.
GN ADRB1 OR ADRB1R.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 90330633.
RA MACHIDA C.A., BUNZOW J.R., SEARLES R.P., VAN TOL H.H.M., TESTER B.,
RA NEVE K.A., TEAL P., NIPPER V., CIVELLI O.;
RT "Molecular cloning and expression of the rat beta 1-adrenergic
receptor gene.";
RL J. BIOL. CHEM. 265:12960-12965(1990).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 90356399.
RA SHIMOMURA H., TERADA A.;
RT "Primary structure of the rat beta-1 adrenergic receptor gene.";
RL NUCLEIC ACIDS RES. 18:4591-4591(1990).
CC -1- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
PROTEINS. THIS RECEPTOR BINDS EPINEPHRINE AND NOREPINEPHRINE WITH
APPROXIMATIVELY EQUAL AFFINITY.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: HOMOLOGOUS DESENSITIZATION OF THE RECEPTOR IS MEDIATED BY ITS
PHOSPHORYLATION BY BETA-ADRENERGIC RECEPTOR KINASE.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; J05561; G203074; -
DR EMBL; D00834; D1000982; -
DR PIR; A36618; A36618.
DR PIR; S12591; S12591.
DR GCRDB; GCR_0126; -
DR GCRDB; GCR_0127; -
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR PFAM; PF00001; 7tm_1; 1.
DR HSP; P07700; 1DEP.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT DOMAIN 1 59 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 60 83 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 84 96 CYTOPLASMIC (POTENTIAL).
FT TRANSNM 97 120 EXTRACELLULAR (POTENTIAL).
FT TRANSNM 121 131 EXTRACELLULAR (POTENTIAL).
FT TRANSNM 132 155 CYTOPLASMIC (POTENTIAL).
FT TRANSNM 156 175 CYTOPLASMIC (POTENTIAL).
FT TRANSNM 176 199 4 (POTENTIAL).
FT TRANSNM 200 221 5 (POTENTIAL).
FT TRANSNM 222 245 5 (POTENTIAL).
FT TRANSNM 246 314 CYTOPLASMIC (POTENTIAL).

Query Match      58.8%; Score 57; DB 1; Length 466;
Best Local Similarity 62.5%; Pred. No. 3.26e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 125 EYGSFFCE 132
QY 575 QYGHYFCE 582

RESULT 7
ID BLAR_MOUSE STANDARD; PRT; 466 AA.
AC P34971;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE BETA-1 ADRENERGIC RECEPTOR.
GN ADRB1 OR ADRB1R.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
[1]
RN SEQUENCE FROM N.A.
RX STRAIN; 129/SV.
RX MEDLINE; 93372116.
RA JASPER J.R., LINK R.E., CHRUSCINSKI A.J., KOBILKA B.K., BERNSTEIN D.;
RT "Primary structure of the mouse beta 1-adrenergic receptor gene.";
RL BIOCHIM. BIOPHYS. ACTA 1178:307-309(1993).
CC -1- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
PROTEINS. THIS RECEPTOR BINDS EPINEPHRINE AND NOREPINEPHRINE WITH
APPROXIMATIVELY EQUAL AFFINITY.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: HOMOLOGOUS DESENSITIZATION OF THE RECEPTOR IS MEDIATED BY ITS
PHOSPHORYLATION BY BETA-ADRENERGIC RECEPTOR KINASE.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; L10084; G293279; -
DR PIR; S36794; S36794.
DR GCRDB; GCR_0578; -
DR MGD; MGI:87937; ADRB1.
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR PFAM; PF00001; 7tm_1; 1.
DR HSP; P07700; 1DEP.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT DOMAIN 1 59 EXTRACELLULAR (POTENTIAL).
FT TRANSNM 60 83 1 (POTENTIAL).
FT TRANSNM 84 96 CYTOPLASMIC (POTENTIAL).
FT TRANSNM 97 120 2 (POTENTIAL).
FT TRANSNM 121 131 EXTRACELLULAR (POTENTIAL).
FT TRANSNM 132 155 3 (POTENTIAL).
```

FT DOMAIN 156 175 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 176 199 4 (POTENTIAL).
 FT DOMAIN 200 221 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 222 245 5 (POTENTIAL).
 FT DOMAIN 246 314 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 315 338 6 (POTENTIAL).
 FT DOMAIN 339 345 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 346 369 7 (POTENTIAL).
 FT DOMAIN 370 466 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 15 15 PROBABLE.
 FT DISULFID 131 209 BY SIMILARITY.
 FT LIPID 381 381 PALMITATE (BY SIMILARITY).
 FT MOD_RES 296 296 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
 FT MOD_RES 301 301 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
 FT MOD_RES 401 401 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
 SQ SEQUENCE 466 AA; 50479 MW; DE499A68 CRC32;

Query Match 58.8%; Score 57; DB 1; Length 466;
 Best Local Similarity 62.5%; Pred. No. 3.26e+00;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 125 EYGSFCE 132
 :|| :|||
 Qy 575 QYGHYFCE 582

RESULT 8
 ID BIAR_PIG STANDARD; PRT; 468 AA.
 AC Q28998; Q46575;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE BETA-1 ADRENERGIC RECEPTOR.
 GN ADRB1.
 OS SUS SCROFA (PIG).
 OC EUKARYOTA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA CAO H., BIDWELL C.A., WILLIAMS S.K., LIANG W., MILLS S.E.;
 RT "Nucleotide sequence of the porcine B1-adrenergic receptor gene."
 RL J. ANIM. SCI. 0:0-0(1998).
 RN [2]
 RP SEQUENCE OF 101-468 FROM N.A.
 RC TISSUE=HEART;
 RA MCNEEL R.L., MERSMANN H.J.;
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
 CC INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
 CC PROTEINS. THIS RECEPTOR BINDS EPINEPHRINE AND NOREPINEPHRINE WITH
 CC APPROXIMATIVELY EQUAL AFFINITY (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- PTM: HOMOLOGOUS DESENSITIZATION OF THE RECEPTOR IS MEDIATED BY ITS
 CC PHOSPHORYLATION BY BETA-ADRENERGIC RECEPTOR KINASE (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF042454; G2796173; -
 DR EMBL; U56425; G2984633; -
 DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; PARTIAL.
 DR PFAM; PF00001; 7tm.1; 1.
 DR HSP; P07700; 10EP.
 KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
 FT MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
 DOMAIN 1 59 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 60 83 1 (POTENTIAL).
 FT DOMAIN 84 96 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 97 121 2 (POTENTIAL).
 FT DOMAIN 122 132 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 133 153 3 (POTENTIAL).
 FT DOMAIN 154 176 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 177 197 4 (POTENTIAL).
 FT DOMAIN 198 223 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 224 244 5 (POTENTIAL).
 FT DOMAIN 245 315 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 316 336 6 (POTENTIAL).
 FT DOMAIN 337 347 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 348 368 7 (POTENTIAL).
 FT DOMAIN 369 468 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 15 15 BY SIMILARITY.
 FT DISULFID 131 208 PALMITATE (BY SIMILARITY).
 FT LIPID 382 382 A -> AR (IN REF. 2).
 FT CONFLICT 173 173 L -> V (IN REF. 2).
 FT CONFLICT 316 316 CWL -> WWV (IN REF. 2).
 FT CONFLICT 326 328 R -> A (IN REF. 2).
 FT CONFLICT 448 448 T -> S (IN REF. 2).
 FT CONFLICT 458 458 AS -> SF (IN REF. 2).
 FT CONFLICT 463 464
 SQ SEQUENCE 468 AA; 50098 MW; E439B14A CRC32;

Query Match 58.8%; Score 57; DB 1; Length 468;
 Best Local Similarity 62.5%; Pred. No. 3.26e+00;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 125 EYGSFCE 132
 :|| :|||
 Qy 575 QYGHYFCE 582

RESULT 9
 ID BIAR_HUMAN STANDARD; PRT; 477 AA.
 AC P08588;
 DT 01-AUG-1988 (REL. 08, CREATED)
 DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE BETA-1 ADRENERGIC RECEPTOR.
 GN ADRB1 OR ADRB1R OR BIAR.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA FRIELLE T., COLLINS S., DANIEL K.W., CARON M.G., LEFKOWITZ R.J.,
 RA KOBILKA B.K.;
 RT "Cloning of the cDNA for the human beta 1-adrenergic receptor."
 RL PROC. NATL. ACAD. SCI. U.S.A. 84:7920-7924(1987).
 CC -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
 CC INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
 CC PROTEINS. THIS RECEPTOR BINDS EPINEPHRINE AND NOREPINEPHRINE WITH
 CC APPROXIMATIVELY EQUAL AFFINITY.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- PTM: HOMOLOGOUS DESENSITIZATION OF THE RECEPTOR IS MEDIATED BY ITS
 CC PHOSPHORYLATION BY BETA-ADRENERGIC RECEPTOR KINASE.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J03019; G178200; -
 DR EMBL; A39911; A39911.
 DR GCRDB; GCR_0048; -
 DR MIN; 109630; -

DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
DR PFAM; PF00001; 7tm_1; 1.
DR HSP; P07700; 1DEP.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT DOMAIN 1 59
FT TRANSMEM 60 83
FT DOMAIN 84 96
FT TRANSMEM 97 120
FT DOMAIN 121 131
FT TRANSMEM 132 155
FT DOMAIN 156 175
FT TRANSMEM 176 199
FT DOMAIN 200 221
FT TRANSMEM 222 245
FT DOMAIN 246 325
FT TRANSMEM 326 349
FT DOMAIN 350 356
FT TRANSMEM 357 380
FT DOMAIN 381 477
FT CARBOHYD 15 15
FT DISULFID 131 209
FT MOD_RES 312 312
FT MOD_RES 412 412
FT LIPID 392 392
SQ SEQUENCE 477 AA; 51223 MW; C1636966 CRC32;

Query Match 58.8%; Score 57; DB 1; Length 477;
Best Local Similarity 62.5%; Pred. No. 3.26e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 125 EYGSFCE 132
QY 575 QYGHYFCE 582

RESULT 10
ID BLAR_MACMU STANDARD; PRT; 480 AA.
AC P47899;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE BETA-1 ADRENERGIC RECEPTOR.
GN ADRB1.
OS MACACA MULATTA (RHESUS MACAQUE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95078456.
RA SEARLES R.P., NIPPER V.J., MACHIDA C.A.;
RT "The rhesus macaque beta 1-adrenergic receptor gene: structure of the
RT gene and comparison of the flanking sequences with the rat beta 1-
RT adrenergic receptor gene."
RL DNA SEQ. 4:231-241(1994).
CC -1- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
CC INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
CC PROTEINS. THIS RECEPTOR BINDS EPINEPHRINE AND NOREPINEPHRINE WITH
CC APPROXIMATIVELY EQUAL AFFINITY.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PFM: HOMOLOGOUS DESENSITIZATION OF THE RECEPTOR IS MEDIATED BY ITS
CC PHOSPHORYLATION BY BETA-ADRENERGIC RECEPTOR KINASE.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X75540; G510533; .

DR GCRDB; GCR_1746; .
DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
DR PFAM; PF00001; 7tm_1; 1.
DR HSP; P07700; 1DEP.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT DOMAIN 1 59
FT TRANSMEM 60 83
FT DOMAIN 84 96
FT TRANSMEM 97 120
FT DOMAIN 121 131
FT TRANSMEM 132 155
FT DOMAIN 156 175
FT TRANSMEM 176 199
FT DOMAIN 200 221
FT TRANSMEM 222 245
FT DOMAIN 246 328
FT TRANSMEM 329 352
FT DOMAIN 353 359
FT TRANSMEM 360 383
FT DOMAIN 384 480
FT CARBOHYD 15 15
FT DISULFID 131 209
FT MOD_RES 315 315
FT MOD_RES 415 415
FT LIPID 395 395
SQ SEQUENCE 480 AA; 51608 MW; 807D224D CRC32;

Query Match 58.8%; Score 57; DB 1; Length 480;
Best Local Similarity 62.5%; Pred. No. 3.26e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 125 EYGSFCE 132
QY 575 QYGHYFCE 582

RESULT 11
ID SYW_SYNV3 STANDARD; PRT; 377 AA.
AC P73555;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRYPTOPHANYL-TRNA SYNTHETASE (EC 6.1.1.2) (TRYPTOPHAN--TRNA LIGASE)
DE (TRPS).
GN TRPS OR SLR1884.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA OKUMURA T., MATSUO A., MURAKI A., NAKAZAKI N., NARUO K.,
RA KOSUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,
RA YAMADA M., YASUDA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA RES. 3:109-136(1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-TRYPTOPHAN + TRNA(Trp) --> AMP +
CC PYROPHOSPHATE + L-TRYPTOPHANYL-TRNA(Trp).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOCYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC


```
CC EMBL: D90908; G1552781; -
DR PROSITE: PS00178; AA:TRNA_LIGASE_I; 1.
DR PFAM: PF00379; TRNA-Sync1b; 1.
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
FT SIMILAR 52 60 "HIGH" REGION.
FT SIMILAR 239 243 "KMSKS" REGION.
FT BINDING 242 242 ATP (BY SIMILARITY).
SQ SEQUENCE 377 AA; 42367 MW; A78606AA CRC32;

Query Match 57.7%; Score 56; DB 1; Length 377;
Best Local Similarity 50.0%; Pred. No. 4.89e+00;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 73 QHYDNFFC 80
QY 574 QQYGHYFC 581
|:|:|:|

RESULT 12
ID BM8A_MOUSE STANDARD; PRT; 399 AA.
AC P34821;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE BONE MORPHOGENETIC PROTEIN 8A PRECURSOR (BMP-8A) (OSTEOGENIC PROTEIN
DE 2) (OP-2).
DE GN BMP8A OR BMP-8.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE; 93094231.
RA OEZKAYNAK E., SCHNEGELSBERG P.N.J., JIN D.F., CLIFFORD G.M.,
RA WARREN F.D., DRIER E.A., OPPERMANN H.;
RT "Osteogenic protein-2. A new member of the transforming growth
RT factor-beta superfamily expressed early in embryogenesis.";
RL J. BIOL. CHEM. 267:25220-25227(1992).
CC -!- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION. MAY BE THE
CC OSTEOINDUCTIVE FACTOR RESPONSIBLE FOR THE PHENOMENON OF
CC EPITHELIAL OSTEOGENESIS. PLAYS A ROLE IN CALCIUM REGULATION
CC AND BONE HOMEOSTASIS (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
CC -!- DEVELOPMENTAL STAGE: EXTENSIVE EXPRESSION FOUND IN 8-DAY EMBRYOS,
CC FELL DRAMATICALLY IN 10-DAY EMBRYOS AND VIRTUALLY ABSENT IN 17-
CC DAY EMBRYOS.
CC -!- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M97017; G200141; -
CC MGD; MGI:104515; BMP8A.
CC PROSITE: PS00250; TGF_BETA; 1.
CC PFAM: PF00019; TGF-beta; 1.
CC PFAM: PF00688; TGFb_propeptide; 1.
CC HSSP: P18075; 1BMP.
CC SIGNAL; GROWTH FACTOR; CYTOKINE; BONE; CARTILAGE; GLYCOPROTEIN.
CC FT SIGNAL 1 19 POTENTIAL.
CC FT PROPEP 20 260 POTENTIAL.
CC FT CHAIN 261 399 BONE MORPHOGENETIC PROTEIN 8A.
CC FT DISULFID 298 364 BY SIMILARITY.
CC FT DISULFID 327 396 BY SIMILARITY.
CC FT DISULFID 331 398 BY SIMILARITY.
CC FT DISULFID 363 363 INTERCHAIN (BY SIMILARITY).
CC FT CARBOHYD 155 155 POTENTIAL.

CC EMBL: M97016; G189390; -
CC PIR; A45056; A45056.
CC MIM; 602284; -
CC PROSITE: PS00250; TGF_BETA; 1.
CC PFAM: PF00019; TGF-beta; 1.
CC PFAM: PF00688; TGFb_propeptide; 1.
CC HSSP: P18075; 1BMP.
CC SIGNAL; GROWTH FACTOR; CYTOKINE; BONE; CARTILAGE; GLYCOPROTEIN.
CC FT SIGNAL 1 19 POTENTIAL.
CC FT PROPEP 20 263 POTENTIAL.
CC FT CHAIN 264 402 BONE MORPHOGENETIC PROTEIN 8.
CC FT DISULFID 301 367 BY SIMILARITY.
CC FT DISULFID 330 399 BY SIMILARITY.
CC FT DISULFID 334 401 BY SIMILARITY.
CC FT DISULFID 366 366 INTERCHAIN (BY SIMILARITY).
CC FT CARBOHYD 158 158 POTENTIAL.
CC FT CARBOHYD 343 343 POTENTIAL.
CC SEQUENCE 402 AA; 44764 MW; 8F20C81B CRC32;

Query Match 57.7%; Score 56; DB 1; Length 402;
Best Local Similarity 41.7%; Pred. No. 4.89e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 321 APOGYSAYYCEG 332
QY 574 QQYGHYFCEA 583
|:|:|:|:|
```

```
FT CARBOHYD 340 340 POTENTIAL.
SQ SEQUENCE 399 AA; 44764 MW; 7075A1ED CRC32;

Query Match 57.7%; Score 56; DB 1; Length 399;
Best Local Similarity 41.7%; Pred. No. 4.89e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 318 APOGYSAYYCEG 329
QY 574 QQYGHYFCEA 583
|:|:|:|:|

RESULT 13
ID BMP8_HUMAN STANDARD; PRT; 402 AA.
AC P34820;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE BONE MORPHOGENETIC PROTEIN 8 PRECURSOR (BMP-8) (OSTEOGENIC PROTEIN 2)
DE (OP-2).
DE GN BMP8.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HIPPOCAMPUS;
RX MEDLINE; 93094231.
RA OEZKAYNAK E., SCHNEGELSBERG P.N.J., JIN D.F., CLIFFORD G.M.,
RA WARREN F.D., DRIER E.A., OPPERMANN H.;
RT "Osteogenic protein-2. A new member of the transforming growth
RT factor-beta superfamily expressed early in embryogenesis.";
RL J. BIOL. CHEM. 267:25220-25227(1992).
CC -!- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION. MAY BE THE
CC OSTEOINDUCTIVE FACTOR RESPONSIBLE FOR THE PHENOMENON OF
CC EPITHELIAL OSTEOGENESIS. PLAYS A ROLE IN CALCIUM REGULATION
CC AND BONE HOMEOSTASIS (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M97016; G189390; -
CC PIR; A45056; A45056.
CC MIM; 602284; -
CC PROSITE: PS00250; TGF_BETA; 1.
CC PFAM: PF00019; TGF-beta; 1.
CC PFAM: PF00688; TGFb_propeptide; 1.
CC HSSP: P18075; 1BMP.
CC SIGNAL; GROWTH FACTOR; CYTOKINE; BONE; CARTILAGE; GLYCOPROTEIN.
CC FT SIGNAL 1 19 POTENTIAL.
CC FT PROPEP 20 263 POTENTIAL.
CC FT CHAIN 264 402 BONE MORPHOGENETIC PROTEIN 8.
CC FT DISULFID 301 367 BY SIMILARITY.
CC FT DISULFID 330 399 BY SIMILARITY.
CC FT DISULFID 334 401 BY SIMILARITY.
CC FT DISULFID 366 366 INTERCHAIN (BY SIMILARITY).
CC FT CARBOHYD 158 158 POTENTIAL.
CC FT CARBOHYD 343 343 POTENTIAL.
CC SEQUENCE 402 AA; 44764 MW; 8F20C81B CRC32;

Query Match 57.7%; Score 56; DB 1; Length 402;
Best Local Similarity 41.7%; Pred. No. 4.89e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 321 APOGYSAYYCEG 332
QY 574 QQYGHYFCEA 583
|:|:|:|:|
```

QY 572 TAQYGHYFCEA 583

RESULT 14

ID FPRP_RAT STANDARD; PRT; 879 AA.
AC Q62786;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROTAGLANDIN F2-ALPHA RECEPTOR REGULATORY PROTEIN PRECURSOR
DE (PROTAGLANDIN F2-ALPHA RECEPTOR ASSOCIATED PROTEIN).
GN FPRP.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:SPRAGUE-DAWLEY; TISSUE-OVARY;
RX MEDLINE; 97109715.
RA ORLICKY D.J., NORDEEN S.K.;
RT "Cloning, sequencing and proposed structure for a prostaglandin F2
alpha receptor regulatory protein."
RL PROTAGLANDINS LEUKOT. ESSENT. FATTY ACIDS 55:261-268(1996).
RN [2]

CHARACTERIZATION.

RC STRAIN:SPRAGUE-DAWLEY; TISSUE-OVARY;
RX MEDLINE; 96397016.
RA ORLICKY D.J.;
RT "Negative regulatory activity of a prostaglandin F2 alpha receptor
associated protein (FPRP).";
RL PROTAGLANDINS LEUKOT. ESSENT. FATTY ACIDS 54:247-259(1996).
CC -!- FUNCTION: INHIBITS THE BINDING OF PROTAGLANDIN F2-ALPHA (PGF2-
ALPHA) TO ITS SPECIFIC FP RECEPTOR. BY DECREASING THE RECEPTOR
NUMBER RATHER THAN THE AFFINITY CONSTANT. FUNCTIONAL COUPLING WITH
THE PROTAGLANDIN F2-ALPHA RECEPTOR SEEMS TO OCCUR.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENDOPLASMIC
RETICULUM AND TRANS-GOLGI NETWORK.
CC -!- TISSUE SPECIFICITY: REPRODUCTIVE TISSUES, LUNG AND HEART.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
6 C2-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

DR EMBL; U26595; G1054884; .
DR PFAM; PF00047; Ig: 6.
KW SIGNAL; IMMUNOGLOBULIN FOLD; TRANSMEMBRANE; GLYCOPROTEIN.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 879 PROTAGLANDIN F2-ALPHA RECEPTOR
FT EXTRACELLULAR (POTENTIAL).
FT DOMAIN 22 832 POTENTIAL.
FT TRANSMEM 833 852 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 854 879 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 36 126 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 162 254 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 292 380 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 422 522 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 564 662 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 704 800 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 424 427 ENDOPLASMIC RETICULAR RETENTION SIGNAL.
FT SITE 89 91 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 703 705 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 43 119 BY SIMILARITY.
FT DISULFID 169 247 BY SIMILARITY.
FT DISULFID 299 373 BY SIMILARITY.
FT DISULFID 429 515 BY SIMILARITY.
FT DISULFID 571 655 BY SIMILARITY.
FT DISULFID 711 793 BY SIMILARITY.

FT CARBOHYD 44 44 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 383 383 POTENTIAL.
FT CARBOHYD 413 413 POTENTIAL.
FT CARBOHYD 525 525 POTENTIAL.
FT CARBOHYD 600 600 POTENTIAL.
FT CARBOHYD 618 618 POTENTIAL.
FT CARBOHYD 691 691 POTENTIAL.
FT VARIANT 47 47 D->G.
FT VARIANT 136 136 M->V.
FT VARIANT 782 782 S->G.
FT VARIANT 844 844 G->R.
SQ SEQUENCE 879 AA; 98730 MW; 0CAEIEEE CRC32;

Query Match 57.7%; Score 56; DB 1; Length 879;
Best Local Similarity 50.0%; Pred. No. 4.89e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 786 QDGNVYC 793

|:|:|:|

QY 574 QYGHYFC 581

RESULT 15

ID PLAS_CHLFU STANDARD; PRT; 98 AA.
AC P00300;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PLASTOCYANIN.
GN PETE.
OS CHLORELLA FUSCA.
OC EUKARYOTA; VIRIDIPANTAE; CHLOROPHYTA; CHLOROPHYCEAE; CHLOROCOCCALES;
OC SCENEDESMACEAE; SCENEDESMUS.
RN SEQUENCE.
RX MEDLINE; 75183336.
RA KELLY J., AMBLER R.P.;
RT "The amino acid sequence of plastocyanin from Chlorella fusca."
RL BIOCHEM. J. 143:681-690(1974).
CC -!- FUNCTION: PLASTOCYANIN PARTICIPATES IN ELECTRON TRANSFER BETWEEN
P700 AND THE CYTOCHROME B/F COMPLEX IN PHOTOSYSTEM I.
CC -!- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
MEMBRANE SURFACE IN CHLOROPLASTS.
CC -!- SIMILARITY: CONTAINS ONE PLASTOCYANIN-LIKE DOMAIN.
DR PIR; A00310; CUKLFC.
DR PROSITE; PS00196; COPPER_BLUE; 1.
DR PFAM; PF00127; copper-bind; 1.
DR MENDEL; I1623; CHLFu; Pete; 1.
DR HSPP; P18068; 2PLT.
DR CHLOROPLAST; ELECTRON TRANSPORT; COPPER; THYLAKOID MEMBRANE.
KW DOMAIN 1 98 PLASTOCYANIN-LIKE.
FT METAL 38 38 COPPER (BY SIMILARITY).
FT METAL 83 83 COPPER (BY SIMILARITY).
FT METAL 86 86 COPPER (BY SIMILARITY).
FT METAL 91 91 COPPER (BY SIMILARITY).
SQ SEQUENCE 98 AA; 10299 MW; D9F2CB91 CRC32;

Query Match 56.7%; Score 55; DB 1; Length 98;
Best Local Similarity 66.7%; Pred. No. 7.30e+00;
Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Db 75 TAGTYG-YFCEP 85

||| ||| |||

QY 572 TAQYGHYFCEA 583

Search completed: Thu Jul 8 19:06:13 1999
Job time : 13 secs.

WIRE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:07:01 1999; Maspar time 5.32 seconds
Tabular output not generated. 123.135 Million cell updates/sec

Title: >US-09-041-236-2
Description: (572-583) from US09041236.pep (29 of 45)
Perfect Score: 97
Sequence: 1 TAQXYGHYFCEA 12

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprembl9

1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertibrate 14:sp.virus

Statistics: Mean 24.892; Variance 37.391; scale 0.666

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	97	100.0	666	4	SEMAPHORIN L.	5.12e-08
2	60	61.9	174	2	HYPOTHETICAL 18.6 KD P	1.76e+00
3	60	61.9	349	2	PUTATIVE GLCNAC TRANSF	1.76e+00
4	59	60.8	376	5	SIMILAR TO UBIQUITIN T	2.66e+00
5	59	60.8	610	5	C44H9.6 PROTEIN.	2.66e+00
6	58	59.8	260	14	GENOME, PARTIAL SEQUEN	4.01e+00
7	58	59.8	565	3	FISSION YEAST (FRAGMEN	4.01e+00
8	58	59.8	641	3	METHYLENETETRAHYDROFOL	4.01e+00
9	58	59.8	884	5	CODED FOR BY C. ELEGAN	4.01e+00
10	57	58.8	157	11	PUTATIVE OLFACTORY REC	6.02e+00
11	57	58.8	276	3	HYPOTHETICAL PROTEIN C	6.02e+00
12	57	58.8	280	2	PUTATIVE HYDROLASE.	6.02e+00
13	57	58.8	909	10	HOMOSERINE DEHYDROGENA	6.02e+00
14	57	58.8	916	8	HOMOSERINE DEHYDROGENA	6.02e+00
15	56	57.7	190	6	PROBABLE PROTAGLANDIN	8.99e+00
16	56	57.7	502	5	MEC-9 PROTEIN.	8.99e+00
17	56	57.7	598	5	F41B5.6 PROTEIN.	8.99e+00
18	56	57.7	838	5	MEC-9L.	8.99e+00
19	56	57.7	838	5	MEC-9 PROTEIN.	8.99e+00
20	56	57.7	1736	10	HYPOTHETICAL 189.0 KD	8.99e+00

21	55	56.7	157	11	063596	PUTATIVE OLFACTORY REC	1.34e+01
22	55	56.7	164	5	Q23640	ZK856_5 PROTEIN.	1.34e+01
23	55	56.7	309	11	063395	OLFACTORY RECEPTOR.	1.34e+01
24	55	56.7	342	2	083194	HYPOTHETICAL 39.2 KD P	1.34e+01
25	55	56.7	470	11	P97842	5-HYDROXYTRYPTAMINE7 R	1.34e+01
26	55	56.7	479	4	P78336	5-HYDROXYTRYPTAMINE7 R	1.34e+01
27	55	56.7	990	5	076351	DNA POLYMERASE ALPHA (1.34e+01
28	55	56.7	1013	2	048436	HSDR.	1.34e+01
29	54	55.7	198	4	Q15877	VOLTAGE-GATED CALCIUM	1.98e+01
30	54	55.7	204	10	065084	PROBABLE PROTEASOME SU	1.98e+01
31	54	55.7	256	14	066182	POLYPROTEIN (FRAGMENTS	1.98e+01
32	54	55.7	275	3	P78853	FISSION YEAST (FRAGMEN	1.98e+01
33	54	55.7	350	7	002869	CHLAMP, G11-ISOFORM PR	1.98e+01
34	54	55.7	512	14	069059	HERPESVIRUS 6 IMMEDIAT	1.98e+01
35	54	55.7	1030	4	076039	STK9 PROTEIN.	1.98e+01
36	54	55.7	1052	10	049676	HYPOTHETICAL 116.4 KD	1.98e+01
37	54	55.7	2127	4	Q13927	VOLTAGE-DEPENDENT L-TY	1.98e+01
38	54	55.7	2138	4	Q13928	VOLTAGE-DEPENDENT L-TY	1.98e+01
39	54	55.7	2146	4	Q13925	VOLTAGE-DEPENDENT L-TY	1.98e+01
40	54	55.7	2155	4	Q13921	VOLTAGE-DEPENDENT L-TY	1.98e+01
41	54	55.7	2157	4	Q99875	VOLTAGE-DEPENDENT L-TY	1.98e+01
42	54	55.7	2158	4	Q13919	CACB RECEPTOR.	1.98e+01
43	54	55.7	2166	6	Q28676	VOLTAGE-DEPENDENT L-TY	1.98e+01
44	54	55.7	2173	4	Q13930	VOLTAGE-DEPENDENT L-TY	1.98e+01
45	54	55.7	3898	14	O92364	POLYPROTEIN.	1.98e+01

ALIGNMENTS

RESULT	1				
ID	075326	PRELIMINARY;	PRT;	666	AA.
AC	075326;				
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L.				
GN	SEMAL.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;				
OC	CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98389619.				
RA	LANG E C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;				
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA				
RT	viruses."				
RL	GENOMICS 51:340-350(1998).				
DR	ENBL; AF030698; G3523115; "				
SQ	SEQUENCE 666 AA; 74823 MW; F546839D CRC32;				

Query Match 100.0%; Score 97; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 5.12e-08;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	604	TAQXYGHYFCEA	615
QY	572	TAQXYGHYFCEA	583

RESULT	2				
ID	053746	PRELIMINARY;	PRT;	174	AA.
AC	053746;				
DT	01-JUN-1998	(TREMBLREL. 06, CREATED)			
DT	01-JUN-1998	(TREMBLREL. 06, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	HYPOTHETICAL 18.6 KD PROTEIN.				
GN	MYO38.05.				
OS	MYCOBACTERIUM TUBERCULOSIS.				
OC	BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;				
OC	ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				

RA BROWN D., CHURCHER C.M.;
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE; 96181548.
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RL "An integrated map of the genome of the tubercle bacillus,
 RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
 RL leprae";
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
 DR EMBL; AL021933; E1252531; -;
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 174 AA; 18622 MW; E9A04078 CRC32;

Query Match 61.9%; Score 60; DB 2; Length 174;
 Best Local Similarity 54.5%; Pred. No. 1.76e+00;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 152 AAPLYGYGYCE 162
 : | | | | | | | | | |
 QY 572 TAQYGYHYFCE 582

RESULT 3
 ID O30629 PRELIMINARY; PRT; 349 AA.
 AC O30629;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE PUTATIVE GLCNAC TRANSFERASE.
 OS ESCHERICHIA COLI.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC ESCHERICHIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-2775 (08:K40);
 RX MEDLINE; 98043545.
 RA AMOR P.A., WHITEFIELD C.;
 RT "Molecular and functional analysis of genes required for expression
 RT of group IB K antigens in Escherichia coli: characterization of the
 RT his-region containing gene clusters for multiple cell-surface
 RT polysaccharides";
 RL MOL. MICROBIOL. 26:145-161(1997).
 DR EMBL; AF013583; G2586167; -;
 DR PFAM; PF00534; Glycos_transf_1; 1.
 KW TRANSFERASE.
 SQ SEQUENCE 349 AA; 40402 MW; 87317E9C CRC32;

Query Match 61.9%; Score 60; DB 2; Length 349;
 Best Local Similarity 60.0%; Pred. No. 1.76e+00;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 139 SQQYAKYLCE 148
 : | | | | | | | | | |
 QY 573 AQQYGYHYFCE 582

RESULT 4
 ID Q94301 PRELIMINARY; PRT; 376 AA.
 AC Q94301;
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE SIMILAR TO UBIQUITIN THIOLESTERASES.
 GN T22F3.2.

OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA WU X.;
 RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA WATERSTON R.;
 RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U64844; G1658342; -;
 DR PFAM; PF00443; UCH-2; 1.
 SQ SEQUENCE 376 AA; 42460 MW; 1E12AFC9 CRC32;

Query Match 60.8%; Score 59; DB 5; Length 376;
 Best Local Similarity 58.3%; Pred. No. 2.56e+00;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 315 TTLDYGHYWCVA 326
 : | | | | | | | | | |
 QY 572 TAQYGYHYFCE 583

RESULT 5
 ID Q18631 PRELIMINARY; PRT; 610 AA.
 AC Q18631;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE C44H9.6 PROTEIN.
 GN C44H9.6.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA SMYE R.;
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL NATURE 368:32-38(1994).
 DR EMBL; 275529; E1344728; -;
 SQ SEQUENCE 610 AA; 70331 MW; BCF9BD2B CRC32;

```

Query Match      60.8%; Score 59; DB 5; Length 610;
Best Local Similarity 66.7%; Pred. No. 2.66e+00;
Matches      8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 72 TAOQNGHGFCHS 83
    |||| || || |
QY 572 TAOQYGHYFCEA 583

RESULT 6
ID Q84467 PRELIMINARY; PRT; 260 AA.
AC Q84467;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GENOME, PARTIAL SEQUENCE.
GN A147L.
OS PARAMECIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95407089.
RA LI Y., LU Z., BURBANK D.E., KUTISH G.F., ROCK D.L., ETTEN J.L.;
RT "Analysis of 43 kb of the Chlorella virus PBCV-1 330-kb genome: map
RT positions 45 to 88."
RL VIROLOGY 212:134-150(1995).
DR EMBL; U42580; G1131491; -.
SQ SEQUENCE 260 AA; 29170 MW; E7F53DCB CRC32;

Query Match      59.8%; Score 58; DB 14; Length 260;
Best Local Similarity 83.3%; Pred. No. 4.01e+00;
Matches      5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 211 YSHYFC 216
    |:|||||
QY 576 YGHYFC 581

RESULT 7
ID P78770 PRELIMINARY; PRT; 566 AA.
AC P78770;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-JAN-1998 (TREMBLREL. 09, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE FISSION YEAST (FRAGMENT).
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98162722.
RA YOSHIOKA S., KATO K., NAKAI K., OKAYAMA H., NOJIMA H.;
RT "Identification of open reading frames in Schizosaccharomyces pombe
RT cDNAs."
RL DNA RES. 4:363-369(1997).
DR EMBL; D89118; D1014477; -.
FT NON_TER 1 1
SQ SEQUENCE 566 AA; 63529 MW; 8CD21F9C CRC32;

Query Match      59.8%; Score 58; DB 3; Length 566;
Best Local Similarity 62.5%; Pred. No. 4.01e+00;
Matches      5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 130 EKYGDYFC 137
    : || |||
QY 574 QOYGHYFC 581

RESULT 8
ID O74927 PRELIMINARY; PRT; 641 AA.

```

```

AC O74927;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE METHYLENETETRAHYDROFOLATE REDUCTASE.
GN MTHFR2.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=968;
RA NAULA N.;
RL THESIS (1998), INSTITUTE FOR GENERAL MICROBIOLOGY,
RL UNIVERSITY OF BERNE, BALTZERSTRASSE.
DR EMBL; AJ011686; E1329960; -.
SQ SEQUENCE 641 AA; 72140 MW; 1BE0630E CRC32;

Query Match      59.8%; Score 58; DB 3; Length 641;
Best Local Similarity 62.5%; Pred. No. 4.01e+00;
Matches      5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 140 EKYGDYFC 147
    : || |||
QY 574 QOYGHYFC 581

RESULT 9
ID Q20627 PRELIMINARY; PRT; 884 AA.
AC Q20627;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CODED FOR BY C. ELEGANS CDNA YK261.3.
GN F49E8.3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNITEA; RHABDITIDA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=BRISTOL N2;
RC MURRAY J., WOHLDMANN P.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=BRISTOL N2;
RC WATERSTON R.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U61949; G1397291; -.
DR PFAM; PF00099; zn-protease; 1.
SQ SEQUENCE 884 AA; 100008 MW; DEFEBAB9 CRC32;

Query Match      59.8%; Score 58; DB 5; Length 884;
Best Local Similarity 63.6%; Pred. No. 4.01e+00;
Matches      7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

```

```

Db 785 GQYAWKYFCE 795
   :|||: ||||
QY 573 AQYGY-HYFCE 582

RESULT 10
ID Q63591 PRELIMINARY: PRT: 157 AA.
AC Q63591;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE OLFACTORY RECEPTOR (FRAGMENT).
GN TPCRI2.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RX MEDLINE: 97224452.
RA VANDERHAEGHEN P., SCHURMANS S., VASSART G., PARMENTIER M.;
RT "Specific repertoire of olfactory receptor genes in the male germ
   cells of several mammalian species.";
RL GENOMICS 39:239-246(1997).
DR EMBL: X89700; G902717; -.
DR PFAM: PF00001; 7tm_1; 1.
FT NON_TER 1
FT NON_TER 157
SQ SEQUENCE 157 AA: 17743 MW: 305FFBA8 CRC32;

Query Match 58.8%; Score 57; DB 11; Length 157;
Best Local Similarity 45.8%; Pred. No. 6.02e+00;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 45 SQEIDHDFCE 55
   :||: ||||
QY 572 TAQYGHYFCE 582

RESULT 11
ID Q36033 PRELIMINARY: PRT: 276 AA.
AC Q36033;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN C4F10.21 IN CHROMOSOME 1 (FRAGMENT).
GN SPAC4F10.21.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:972;
RA CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: SOME, TO C.ELEGANS T20B12.1 AND YEAST YNL313C.
DR EMBL: 298980; E339886; -.
KW HYPOTHETICAL PROTEIN.
FT NON_TER 276
FT NON_TER 276
SQ SEQUENCE 276 AA: 32110 MW: E4C66A12 CRC32;

Query Match 58.8%; Score 57; DB 3; Length 276;
Best Local Similarity 30.0%; Pred. No. 6.02e+00;
Matches 3; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 17 STEYDQDFC 26
   :|||: |||
QY 572 TAQYGHYFCE 581

RESULT 12

```

```

ID Q69808 PRELIMINARY: PRT: 280 AA.
AC Q69808;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE HYDROLASE.
GN SCIA6.03.
OS STREPTOMYCES COELICOLOR.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCETACEAE; STREPTOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA OLIVER K., HARRIS D.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97000351.
RA REDENBACH M., KIESER H.M., DENAPAITTE D., EICHNER A., CULLUM J.,
   KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL MOL. MICROBIOL. 21:77-96(1996).
DR EMBL: AL023496; E1292351; -.
KW HYDROLASE.
SQ SEQUENCE 280 AA: 31804 MW: F09ED54A CRC32;

Query Match 58.8%; Score 57; DB 2; Length 280;
Best Local Similarity 71.4%; Pred. No. 6.02e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 47 YGPFYFCQ 53
   ||||:
QY 576 YGHYFCE 582

RESULT 13
ID Q65027 PRELIMINARY: PRT: 909 AA.
AC Q65027;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE HOMOSERINE DEHYDROGENASE (EC 1.1.1.3) (HDH) (FRAGMENT).
GN AK-HSDH.
OS GLYCINE MAX (SOYBEAN).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC FABALES; FABACEAE; PAPILIONOIDEAE; GLYCINE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. CENTURY;
RA GEBHARDT J.S., WEISEMANN J.M., MATTHEWS B.F.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: L-HOMOSERINE + NAD(P)+(-) -> L-ASPARTATE
   BETA-SEMIALDEHYDE + NAD(P)H.
CC -!- PATHWAY: THIRD STEP IN CONVERSION OF L-ASPARTATE TO HOMOSERINE.
CC HOMOSERINE PARTICIPATES IN THE BIOSYNTHESIS OF THREONINE AND THEN
   ISOLEUCINE AND IN THE BIOSYNTHESIS OF METHIONINE.
CC -!- SIMILARITY: TO OTHER SPECIES HOMOSERINE DEHYDROGENASE.
DR EMBL: AF049708; G2970536; -.
DR EMBL: AF049707; G2970536; JOINED.
DR PROSITE: PS00324; ASPARTOKINASE; 1.
DR PROSITE: PS01042; HOMOSER.DHGENASE; 1.
KW OXIDOREDUCTASE; NADP; THREONINE BIOSYNTHESIS; ISOLEUCINE BIOSYNTHESIS;
   METHIONINE BIOSYNTHESIS.
FT NON_TER 909
FT NON_TER 909
SQ SEQUENCE 909 AA: 99723 MW: 6C065934 CRC32;

```

Query Match 58.8%; Score 57; DB 10; Length 909;
Best Local Similarity 70.0%; Pred. No. 6.02e+00;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 692 QSYTHFYFEA 701
| | | | |
QY 574 QQYGHYFCEA 583

RESULT 14
ID O63067; PRELIMINARY; PRT; 916 AA.
AC O63067;
DT 01-AUG-1998 (TRENBLREL. 07, CREATED)
DT 01-AUG-1998 (TRENBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TRENBLREL. 07, LAST ANNOTATION UPDATE)
DE HOMOSERINE DEHYDROGENASE (EC 1.1.1.3) (HSDH).
GN AK-HSDH.
OS GLYCINE MAX (SOYBEAN).
OG CHLOROPLAST.
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC FABACEAE; PAPILIONOIDEAE; GLYCINE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CENTURY;
RA GEBHARDT J.S.; WEISEMANN J.M.; MATTHEWS B.F.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: L-HOMOSERINE + NAD(P)(+) = L-ASPARTATE
BETA-SEMIALDEHYDE + NAD(P)H.
CC -!- PATHWAY: THIRD STEP IN CONVERSION OF L-ASPARTATE TO HOMOSERINE.
CC HOMOSERINE PARTICIPATES IN THE BIOSYNTHESIS OF THREONINE AND THEN
ISOLEUCINE AND IN THE BIOSYNTHESIS OF METHIONINE.
CC -!- SIMILARITY: TO OTHER SPECIES HOMOSERINE DEHYDROGENASE.
DR EMBL; AF049706; G2970447; -
DR PROSITE; PS00324; ASPARTOKINASE; 1.
DR PROSITE; PS01042; HOMOSER. DHGENASE; 1.
KW TRANSFERASE; CHLOROPLAST; OXIDOREDUCTASE; NADP;
KW THREONINE BIOSYNTHESIS; ISOLEUCINE BIOSYNTHESIS;
KW METHIONINE BIOSYNTHESIS.
SQ SEQUENCE 916 AA; 100394 MW; 6D9B5C12 CRC32;

Query Match 58.8%; Score 57; DB 8; Length 916;
Best Local Similarity 70.0%; Pred. No. 6.02e+00;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 692 QSYTHFYFEA 701
| | | | |
QY 574 QQYGHYFCEA 583

RESULT 15
ID O02834; PRELIMINARY; PRT; 190 AA.
AC O02834;
DT 01-JUL-1997 (TRENBLREL. 04, CREATED)
DT 01-JUL-1997 (TRENBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TRENBLREL. 05, LAST ANNOTATION UPDATE)
DE PROBABLE PROSTAGLANDIN F2-ALPHA RECEPTOR REGULATORY PROTEIN
(FRAGMENT).
DE (FRAGMENT).
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RA VERGIN K.L.;
RL THESIS (1997), OREGON STATE UNIVERSITY, IN PRESS.
CC -!- FUNCTION: INHIBITS THE BINDING OF PROSTAGLANDIN F2-ALPHA
(PGF2-ALPHA) TO ITS SPECIFIC FP RECEPTOR, BY DECREASING THE
RECEPTOR NUMBER RATHER THAN THE AFFINITY CONSTANT.
CC FUNCTIONAL COUPLING WITH THE PROSTAGLANDIN F2-ALPHA RECEPTOR
SEEMS TO OCCUR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
DR EMBL; AF004133; G2198826; -

DR EMBL; AF004134; G2198828; -
KW IMMUNOGLOBULIN FOLD; TRANSMEMBRANE; GLYCOPROTEIN.
FT NON_TER 1
FT DOMAIN >1 143 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 144 164 POTENTIAL.
FT DOMAIN 165 190 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 2 2 POTENTIAL.
SQ SEQUENCE 190 AA; 21326 MW; A85F0AFB CRC32;

Query Match 57.7%; Score 56; DB 6; Length 190;
Best Local Similarity 50.0%; Pred. No. 8.99e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 97 QDFGNYYC 104
| | | | |
QY 574 QQYGHYFC 581

Search completed: Thu Jul 8 19:07:20 1999
Job time : 19 secs.

THIS PAGE BLANK (USPTO)

WIREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:09:43 1999; MasPar time 11.10 Seconds
Tabular output not generated. 17.242 Million cell updates/sec

Title: >US-09-041-236-2
Description: (598-606) from US09041236.pep (30 of 45)
Perfect Score: 63
Sequence: 1 LPEDGIMAE 9

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21265608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 15.680; Variance 46.051; scale 0.340

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	47	74.6	147	P81931	Human basic fibroblas	1.06e+02
2	47	74.6	249	W81550	Bacillus subtilis rib	1.06e+02
3	47	74.6	302	W59495	Human FEF3 5'-RACE fu	1.06e+02
4	47	74.6	361	P10070	B.subtilis rib ORF 5	1.06e+02
5	47	74.6	389	W28223	Human reciprocal TFE3	1.06e+02
6	47	74.6	618	W52827	Human chromosome X TF	1.06e+02
7	45	71.4	152	W35486	IL-1 inhibitor (IL-1i	1.76e+02
8	45	71.4	153	W61149	Recombinant human int	1.76e+02
9	45	71.4	153	W22894	Recombinant human int	1.76e+02
10	45	71.4	159	W83008	Intracellular interleu	1.76e+02
11	45	71.4	159	R73642	icIL-1ra.	1.76e+02
12	45	71.4	159	R15262	Variant IL-1 cytokine	1.76e+02
13	45	71.4	159	W80777	Human intracellular I	1.76e+02
14	45	71.4	159	W37787	Recombinant interleuk	1.76e+02
15	45	71.4	165	P93616	Sequence encoded by b	1.76e+02
16	45	71.4	177	R73641	IL-1ra.	1.76e+02

17	45	71.4	177	18	R99263	Human interleukin-1 r	1.76e+02
18	45	71.4	177	18	R98253	Interleukin-1 recepto	1.76e+02
19	45	71.4	177	18	R98252	Interleukin-1 recepto	1.76e+02
20	45	71.4	177	7	R35489	IL-1 inhibitor (IL-1i	1.76e+02
21	45	71.4	177	18	R99261	Human interleukin-1 r	1.76e+02
22	45	71.4	177	4	P91515	Sequence of interleuk	1.76e+02
23	45	71.4	177	30	W56150	Amino acid sequence o	1.76e+02
24	45	71.4	177	3	R14400	IRAP.	1.76e+02
25	45	71.4	177	7	R35485	IL-1i-2A fragment.	1.76e+02
26	45	71.4	177	18	R98251	Interleukin-1 recepto	1.76e+02
27	45	71.4	177	4	P96159	Sequence of interleuk	1.76e+02
28	45	71.4	177	18	R99262	Human interleukin-1 r	1.76e+02
29	45	71.4	177	14	R75784	Human IL-1 receptor a	1.76e+02
30	45	71.4	180	18	R91361	Intracellular IL-1 re	1.76e+02
31	45	71.4	287	20	W06967	Cyclopropane mycolic	1.76e+02
32	45	71.4	1487	13	R74634	Bacillus circulans CF	1.76e+02
33	43	68.3	31	17	R85061	Peptide rII from the	2.89e+02
34	43	68.3	212	31	W56797	S. thermophilus F1 po	2.89e+02
35	43	68.3	325	5	R25074	PLAP.	2.89e+02
36	43	68.3	325	17	R85872	WD-40 domain-contg. p	2.89e+02
37	43	68.3	325	19	R91347	Murine PLAP, for redu	2.89e+02
38	43	68.3	462	37	W73012	Cobra venom protease	2.89e+02
39	43	68.3	467	37	W75057	Human secreted protei	2.89e+02
40	43	68.3	592	37	W73011	Cobra venom protease	2.89e+02
41	43	68.3	1891	26	W23720	Platenolide synthase	2.89e+02
42	43	68.3	1891	26	W22610	Platenolide synthase	2.89e+02
43	42	66.7	1050	39	W67642	A serine/threonine pr	3.70e+02
44	42	66.7	1051	39	W67641	A serine/threonine pr	3.70e+02
45	42	66.7	1928	22	W19664	ATM mutant 5712insA.	3.70e+02

ALIGNMENTS

RESULT 1
ID P81931 standard; protein; 147 AA.
AC P81931:
DE 26-OCF-1990 (first entry)
KW Human basic fibroblast growth factor mutain TM910 from phage M13-PN10.
KW Human basic fibroblast growth factor; human bFGF mutain TM910;
KW growth promoting activity; growth stimulating activity; phage M13-PN10;
KW capillary endothelial cells; angiogenic activity.
OS Synthetic.
PH Key Location/Qualifiers
FT cross_links 9..9
FT /label=mutation_Gly_to_Thr
FT /note="creates NcoI recognition site"
FT misc_difference 10..10
FT /label=mutation_Ser_to_Met
FT /note="creates NcoI recognition site"

EP-281822-A.
PD 14-SEP-1988.
PF 20-FEB-1988; 102491.
PR 24-FEB-1987; JP-042218.
PA (TAKE) Takeda Chemical Ind KK.
PI Senoo M, Krokawa T, Igarashi K, Sasada R;
DR WPI; 88-358590/37.
DR N-PDSB; N81988.
PT Mutain of basic fibroblast growth factor -
PT having fibroblast growth promoting activity, growth stimulating
PT activity of capillary endothelial cells and angiogenic activity..
PS Disclosure: lpp; English.
CC Using plasmid pTB854, E.coli MM294 was transformed, whereby the
CC strain E.coli MM294/pTB854 was obtained, which
CC harbors the plasmid pTB854 expressing the mutain represented here.
CC The mutain has high stability and is low in toxicity.
CC It can be used as a healing accelerator for e.g. burns, wounds
CC or postoperative tissues or as a therapeutic drug based on its
CC angiogenic action for e.g. thrombosis or arteriosclerosis. It can also
CC be used as a reagent for acceleration of cell cultivation. A mutain
CC where at least one constituent cysteine is replaced by serine is
CC preferred because the mutain is highly stable and intermolecular bridges
CC and linkages are reduced or eliminated.
CC See also N81971-97.

```

SQ Sequence 147 AA;
Query Match 74.6%; Score 47; DB 1; Length 147;
Best Local Similarity 75.0%; Pred. No. 1.06e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 4 lpedgtmg 11
    |||||
QY 598 LPEDGIMA 605

RESULT 2
ID W81550 standard; Protein; 249 AA.
AC W81550;
DT 01-MAR-1999 (first entry)
DE Bacillus subtilis rib-specific deaminase.
KW Riboflavin; vitamin B2; deaminase; rib operon.
OS Bacillus subtilis strain 168.
PN US5837528-A.
PD 17-NOV-1998.
PF 06-FEB-1995; 384626.
PR 11-SEP-1990; US-581048.
PR 22-JUN-1989; US-370378.
PR 21-APR-1992; US-873572.
PR 06-FEB-1995; US-384626.
PA (HOFF ) HOFFMANN LA ROCHE INC.
PI Perkins JB, Pero JG, Sloma A;
DR WPI: 99-023456/02.
DR N-PSDB; V71600.
PT Nucleic acid encoding proteins involved in bacterial synthesis of
PT riboflavin - also mutant or recombinant bacteria producing
PT riboflavin at high levels
PS Example 7; Fig 3A-S; 62pp; English.
CC This is the amino acid sequence of a 44 kba rib-specific deaminase
CC encoded by open reading frame 5 (ORF5) of the Bacillus subtilis rib
CC operon (see V71600). The deaminase is predicted to catalyse the
CC formation of 5-amino-6-(ribosylamino)-2,4(1H,3H)-pyrimidinodione
CC 5'-phosphate from 2,5-diamino-6-(ribosylamino)-4(3H)-pyrimidinodione
CC 5-phosphate, and shows 39% identity to the deoxycytidylate
CC deaminase of E. coli phage T2. The invention provides vectors and
CC recombinant bacteria for overproducing riboflavin, in which nucleic
CC acids overproducing riboflavin biosynthetic proteins (see
CC W81549-55) are introduced in the chromosome of the host organism at
CC multiple sites or in multiple copies. The rib operon, e.g.
CC comprising ORFs 2-5 and the beta-riboflavin synthase gene, is used
CC to make such recombinant bacteria. The rib operon can be
CC deregulated by replacing its control regions with sequences that
CC allow constitutive or unregulated expression. The recombinant
CC bacteria produce at least 10 g/l riboflavin.
SQ Sequence 249 AA;

Query Match 74.6%; Score 47; DB 39; Length 249;
Best Local Similarity 62.5%; Pred. No. 1.06e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 167 laeegims 174
    |.:|||:
QY 598 LPEDGIMA 605

RESULT 3
ID W59495 standard; Protein; 302 AA.
AC W59495;
DT 08-SEP-1998 (first entry)
DE Human TFE3 5'-RACE fusion product from cell line UOK109.
KW PRCC; papillary renal cell carcinoma; TFE3; transcription factor; RACE;
KW fusion protein; translocation; diagnosis; treatment; NonO; p54-nrb.
OS Homo sapiens.
OS Synthetic.
FH Kcy Location/Qualifiers
FT Protein 1..302
FT /label="TFE3
FT /note="partial transcription factor sequence"

PN WO9806871-A1.
PD 19-FEB-1998.
PF 13-AUG-1997; G02209.
PR 13-AUG-1996; GB-016986.
PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PI Clark J, Cooper C, Shipley J;
DR WPI: 98-159557/14.
DR N-PSDB; V20961.
PT Diagnosing papillary renal cell carcinoma by detecting gene
PT trans-location - resulting in fusion of TFE3 gene with some other
PT gene, also related vectors, transformed cells, specific binding
PT reagents, peptide(s) encoded by fusions and therapeutic anti-sense
PT sequences
PS Claim 9; Fig 4B; 71pp; English.
CC This sequence represents a partial transcription factor, TFE3 sequence
CC from a 5'-RACE product derived from a novel fusion protein
CC constructed from the NonO protein (also known as p54-nrb) and the
CC transcription factor TFE3 which is found in cell line UOK109 and used in
CC a method for the diagnosis, prophylactic and therapeutic treatment of
CC papillary renal cell carcinoma. The translocation t(X;1) (p11.2;q21.2)
CC found in papillary renal cell carcinoma (PRCC) associated protein (PRCC)
CC results in a fusion of the TFE3 gene with a new chromosome 1 gene
CC designated PRCC (at lq21.2), resulting in expression of a fusion protein
CC between the N-terminus of PRCC and almost the whole of the TFE3 gene.
CC Normal TFE3 transcripts are no longer produced. Two other fusion partners
CC for TFE3 have also been detected; NonO, from a invX (p11.2; q13-24 or 12)
CC translocation and the PSF splice factor gene, resulting in t(X;1)
CC (p11.2;p34). These trans-locations define a subgroup of PRCC generally
CC encountered in patients younger than 25.
SQ Sequence 302 AA;

Query Match 74.6%; Score 47; DB 32; Length 302;
Best Local Similarity 66.7%; Pred. No. 1.06e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 lpsgivid 52
    ||| |||:
QY 598 LPEDGIMAE 606

RESULT 4
ID R10070 standard; Protein; 361 AA.
AC R10070;
DT 19-MAR-1991 (first entry)
DE B.subtilis rib ORF 5 gene product.
KW Riboflavin; vitamin B2.
OS Bacillus subtilis.
PN EP-405370-A.
PD 02-JAN-1991.
PF 22-JUN-1990; 111916.
PR 22-JUN-1989; US-370378.
PA (HOFF ) HOFFMANN-LA ROCHE AG.
PI Perkins JB, Pero JG, Sloma A;
DR WPI: 91-008756/02.
DR Q-PSDB; Q10165.
PT DNA encoding riboflavin biosynthetic proteins - used to produce
PT recombinant bacteria for prodn. of riboflavin
PS Disclosure; Fig 3; 72pp; English.
CC The peptide is a product of the B. subtilis rib operon, involved in
CC the biosynthesis of riboflavin.
CC The product may be expressed in a transformed bacterial expression
CC system for the efficient production of riboflavin (vitamin B2).
SQ Sequence 361 AA;

Query Match 74.6%; Score 47; DB 2; Length 361;
Best Local Similarity 62.5%; Pred. No. 1.06e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 279 laeegims 286
    |.:|||:
QY 598 LPEDGIMA 605

```

RESULT 5
 ID W52823 standard; Protein; 389 AA.
 AC W52823;
 DT 08-SEP-1998 (first entry)
 DE Human reciprocal TFE3/NonO fusion product from cell line UOK109.
 KW PRCC: papillary renal cell carcinoma; TFE3; transcription factor;
 fusion protein; translocation; diagnosis; treatment; NonO; p54-nrb.
 OS Homo sapiens.
 PS Synthetic.
 FH Key
 FT Protein
 FT 1.389
 FT Location/Qualifiers
 FT 1.389
 FT /label= TFE3/NonO
 FT /note= "reciprocal fusion protein"
 FT Misc_difference 1.295
 FT /label= TFE3
 FT /note= "transcription factor sequence"
 FT Misc_difference 296.389
 FT /label= NonO
 FT /note= "also known as p54-nrb"
 WO9806871-A1.
 PN 19-FEB-1998.
 PD 13-AUG-1997; G02209.
 PF 13-AUG-1996; GB-016986.
 PR (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
 PA Clark J. Cooper C, Shipley J;
 PI WPI: 98-159557/14.
 DR N-PSDB: V20959.
 DT Diagnosing papillary renal cell carcinoma by detecting gene
 PT trans-location - resulting in fusion of TFE3 gene with some other
 PT gene, also related vectors, transformed cells, specific binding
 PT reagents, peptide(s) encoded by fusions and therapeutic anti-sense
 PT sequences
 PS Claim 9; Fig 3B; 7lpp; English.
 CC This sequence represents a novel reciprocal fusion protein constructed
 CC from the NonO protein (also known as p54-nrb) and the transcription
 CC factor TFE3 which is found in cell line UOK109 and used in a method for
 CC the diagnosis, prophylactic and therapeutic treatment of papillary renal
 CC cell carcinoma. The translocation t(X;1) (p11.2;q21.2) found in papillary
 CC renal cell carcinoma (PRCC) associated protein (PRCC) results in a fusion
 CC of the TFE3 gene with a new chromosome 1 gene designated PRCC
 CC (at lq21.2), resulting in expression of a fusion protein between the
 CC N-terminus of PRCC and almost the whole of the TFE3 gene. Normal TFE3
 CC transcripts are no longer produced. Two other fusion partners for TFE3
 CC have also been detected; NonO, from a invX (p11.2; q13-24 or 12)
 CC translocation and the PSF splice factor gene, resulting in t(X;1)
 CC (p11.2;p34). These trans-locations define a subgroup of PRCC generally
 CC encountered in patients younger than 25.
 SQ Sequence 389 AA;

Query Match 74.6%; Score 47; DB 32; Length 389;
 Best Local Similarity 66.7%; Pred. No. 1.06e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 lpsgivad 52
 ||| |||:
 QY 598 LPEDGIMAE 606

RESULT 6
 ID W52827 standard; Protein; 618 AA.
 AC W52827;
 DT 08-SEP-1998 (first entry)
 DE Human chromosome X TFE3 protein.
 KW PRCC: papillary renal cell carcinoma; TFE3; transcription factor;
 fusion protein; translocation; diagnosis; treatment; chromosome X.
 OS Homo sapiens.
 PS WO9806871-A1.
 PN 19-FEB-1998.
 PD 13-AUG-1997; G02209.
 PF 13-AUG-1996; GB-016986.
 PR (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
 PA Clark J. Cooper C, Shipley J;
 PI WPI: 98-159557/14.
 DR

DR N-PSDB: V20963.
 PT Diagnosing papillary renal cell carcinoma by detecting gene
 PT trans-location - resulting in fusion of TFE3 gene with some other
 PT gene, also related vectors, transformed cells, specific binding
 PT reagents, peptide(s) encoded by fusions and therapeutic anti-sense
 PT sequences
 PS Claim 13; Fig 5B; 7lpp; English.
 CC This sequence represents the human transcription factor TFE3 which is
 CC used in a method for the diagnosis, prophylactic and therapeutic
 CC treatment of papillary renal cell carcinoma (PRCC). The translocation
 CC t(X;1) (p11.2;q21.2) found in PRCC results in a fusion of the
 CC transcription factor, TFE3 gene with a new chromosome 1 gene designated
 CC PRCC (at lq21.2), resulting in expression of a fusion protein between the
 CC N-terminus of PRCC and almost the whole of the TFE3 gene. Normal TFE3
 CC transcripts are no longer produced. Two other fusion partners for TFE3
 CC have also been detected; NonO, from a invX (p11.2; q13-24 or 12)
 CC translocation and the PSF splice factor gene, resulting in t(X;1)
 CC (p11.2;p34). These trans-locations define a subgroup of PRCC generally
 CC encountered in patients younger than 25.
 SQ Sequence 618 AA;

Query Match 74.6%; Score 47; DB 32; Length 618;
 Best Local Similarity 66.7%; Pred. No. 1.06e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 lpsgivad 52
 ||| |||:
 QY 598 LPEDGIMAE 606

RESULT 7
 ID R35486 standard; Protein; 152 AA.
 AC R35486;
 DT 26-AUG-1993 (first entry)
 DE IL-1 inhibitor (IL-1i).
 KW Interleukin-1 inhibitor; immunosuppressive; inflammation; cytokine;
 KW collagenase; thymocyte; proliferation.
 OS Homo sapiens.
 FH Key
 FT Location/Qualifiers
 FT misc_difference 1
 FT /label= ARG, PRO
 FT /note= "N-terminal Arg is pref."

PN EP-541920-A.
 PD 19-MAY-1993.
 PF 26-MAY-1989; 109540.
 PR 27-MAY-1988; US-199915.
 PR 31-AUG-1988; US-238713.
 PR 23-SEP-1988; US-248521.
 PR 03-NOV-1988; US-266531.
 PR (SYND) SYNERGEN INC.
 PA Arend WP, Eisenberg SP, Hannum CH, Joslin FG, Sommer A;
 PI Thompson RC;
 PI WPI: 93-160536/20.
 DR N-PSDB: Q40754.
 DT New interleukin-1 inhibiting peptide and DNA - useful as
 PT immunosuppressant for treating auto-immune and other immune
 PT disorders
 PS Claim 10; Page 28; 55pp; English.
 CC The IL-1 inhibitor is capable of inhibiting IL-1 induced thymocyte
 CC proliferation.
 CC IL-1i is useful as an immunosuppressive agent. When applied
 CC locally it can be used to prevent tissue destruction in an inflamed
 CC joint and other inflammation sites. This protective effect may be
 CC improved if IL-1i is given with collagenase inhibitors.
 SQ Sequence 152 AA;

Query Match 71.4%; Score 45; DB 7; Length 152;
 Best Local Similarity 42.9%; Pred. No. 1.76e+02;
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 136 mpdegvm 142
 :|||:
 QY 598 LPEDGIM 604

```

CC combination of a rhuIL-lra and a controlled release polymer such as a
CC hydrogel ester (e.g. a hyaluronan or its salt). The receptor antagonist
CC has a relatively short half life, so administering it in a controlled
CC release formulation provides long lasting suppression of interleukin-1
CC (IL-1) activity. The composition is used in human or veterinary medicine
CC to treat diseases mediated by IL-1, such as inflammation, e.g. rheumatoid
CC or osteoarthritis. Conditions resulting from strains, sprains, injury,
CC infection, cartilage damage or orthopaedic surgery can also be treated
CC with this composition. Treatment of a wide range of other diseases is
CC also contemplated, e.g. inflammatory bowel disease, multiple sclerosis,
CC multiple myeloma, myelogenous or other leukemias, brain injuries, graft
CC versus host disease, Alzheimer's disease and temporal mandibular joint
CC disease.
CC Sequence 153 AA;

Query Match 71.4%; Score 45; DB 25; Length 153;
Best Local Similarity 42.9%; Pred. No. 1.76e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 137 mpdegvm 143
Qy 598 LPEDGIM 604
:|::|:|
:|::|:|

RESULT 10
ID W83008 standard; Protein; 159 AA.
AC W83008;
DE 01-FEB-1999 (first entry)
DT Intracellular interleukin 1 receptor antagonist.
KW Intracellular interleukin 1 receptor antagonist; cell adhesion;
KW endothelial cell; intracellular IL-1 receptor antagonist;
KW endometrial cancer; autoimmune related disease;
KW graft versus host disease.
OS Homo sapiens.
PN US840496-A.
PD 24-NOV-1998.
PF 02-JUN-1995; 459814.
PR 08-NOV-1991; US-791474.
PR 09-OCT-1990; US-594827.
PR 18-APR-1995; US-425232.
PR 02-JUN-1995; US-459814.
PA (CHIR ) CHIRON CORP.
PI Haskill JS;
DR N-PSDB; V69830.
PT Diagnosis of endometrial cancer - by measuring intracellular IL-1
PS receptor antagonist level
PS Claim 4; Fig 1; 14pp; English.
CC A method has been developed for the diagnosis of endometrial cancer.
CC The method comprises measuring the amount of intracellular interleukin 1
CC (IL-1) receptor antagonist present in endometrial cells from a patient
CC and comparing the amount with that present in normal endometrial cells.
CC The level of intracellular IL-1 receptor antagonist expression in
CC endometrial cancer cells is significantly higher than in normal cells.
CC The present sequence represents a intracellular IL-1 receptor antagonist.
CC Sequence 159 AA;

Query Match 71.4%; Score 45; DB 37; Length 159;
Best Local Similarity 42.9%; Pred. No. 1.76e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 143 mpdegvm 149
Qy 598 LPEDGIM 604
:|::|:|
:|::|:|

RESULT 11
ID R73642 standard; Protein; 159 AA.
AC R73642;
DE 28-NOV-1995 (first entry)
DT icil-lra.
KW Interleukin-1; receptor antagonist; ovulation prevention; embryo;
KW implantation; contraceptive.

```

OS Homo sapiens.
 PN WO9510298-A.
 PD 20-APR-1995.
 PF 12-OCT-1994; U11588.
 PR 12-OCT-1993; US-136077.
 PA (POLA/) POLAN M L.
 PA (POLA/) POLAN M.
 PA (SIMO/) SIMON C.
 PI Polan ML, Simon C, Polan M;
 DR WPI; 95-169961/22.
 DR N-PSDB; Q89793.
 PT Method of contraception using interleukin-1 receptor antagonist -
 PT prevents ovulation and implantation of embryo, also contraceptive
 PT device comprising the antagonist
 PS Disclosure; Fig 7; 59pp; English.
 CC The sequence is that of the polypeptide iCIL-1ra, an interleukin-1
 CC receptor antagonist. This polypeptide can be used in a
 CC composition (claimed) to prevent ovulation and implantation of an
 CC embryo in a mammalian uterus. The IL-1 receptor antagonist is
 CC useful in contraceptive compans. It is also useful after the
 CC fertilisation event and provide an alternative to known means of
 CC terminating pregnancy post-fertilisation.
 CC See also R73641.
 SQ Sequence 159 AA;

Query Match 71.4%; Score 45; DB 14; Length 159;
 Best Local Similarity 42.9%; Pred. No. 1.76e+02;
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 143 mpdegvm 149
 :|::|:
 QY 598 LPEDGIM 604

RESULT 12
 ID R15262 standard; Protein; 159 AA.
 AC R15262;
 DE 13-FEB-1992 (first entry)
 DT Variant il-1 cytokine inhibitor.
 KW Intracellular; Interleukin-1; cancer; immunosuppressive.
 OS Homo sapiens.
 PN WO9117249-A.
 PD 14-NOV-1991.
 PF 10-APR-1991; U02460.
 PR 01-MAY-1990; US-517276.
 PA (CETU) CETUS CORP.
 PA (UYNC-) UNIV NORTH CAROLINA.
 PI Haskill JS, Martin G, Ralph P;
 DR WPI; 91-353770/48.
 DR N-PSDB; Q14843.
 PT New interleukin-1 antagonists - used to diagnose conditions
 PT mediated by IL-1 and to treat and prevent sepsis and cancer
 PS Claim 7; Fig 2; 42pp; English.
 CC The amino acid sequence is that of an intracellular protein having
 CC cytokine inhibitory activity it is a variant interleukin-1 (IL-1)
 CC cytokine inhibitor. The DNA encoding the inhibitor can be used to
 CC determine the number of copies of the inhibitor gene present per
 CC cell in various types of cancers and so measure the degree of over-
 CC amplification. The inhibitor can be administered to patients at high
 CC risk of developing sepsis or who have already developed it. It may
 CC also have immunosuppressive effects against rheumatoid arthritis.
 SQ Sequence 159 AA;

Query Match 71.4%; Score 45; DB 3; Length 159;
 Best Local Similarity 42.9%; Pred. No. 1.76e+02;
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 143 mpdegvm 149
 :|::|:
 QY 598 LPEDGIM 604

RESULT 13

ID W80777 standard; Protein; 159 AA.
 AC W80777;
 DE Human intracellular IL-1 receptor antagonist.
 KW Intracellular IL-1 receptor antagonist; interleukin-1; IL-1; human;
 KW graft-versus-host disease; endothelial cell damage; vascular disease;
 KW ischemia-related disease; organ transplantation; myocardial infarction;
 KW stroke; autoimmune disease; IL-1 dependent cancer.
 OS Homo sapiens.
 PN US5817306-A.
 PD 06-OCT-1998.
 PF 07-JUN-1995; 484597.
 PR 08-NOV-1991; US-791474.
 PR 09-OCT-1990; US-594827.
 PR 18-APR-1995; US-425232.
 PR 07-JUN-1995; US-484597.
 PA (CHIR) CHIRON CORP.
 PI Haskill JS, Ralph P;
 DR WPI; 98-556335/47.
 DR N-PSDB; V65200.
 PT Treatment of graft-versus-host disease - with interleukin-1 receptor
 PT antagonist polypeptide
 PS Claim 1; Fig 1; 14pp; English.
 CC This represents a human intracellular interleukin-1 (IL-1) receptor
 CC antagonist. The IL-1 receptor antagonist or an IL-1-binding fragment of
 CC the antagonist can be used treating graft-versus-host disease in a human
 CC patient. The polypeptide can be used in methods and medicaments for
 CC treating and preventing diseases that result in endothelial cell damage
 CC such as ischemia-related diseases, including stroke, vascular disease,
 CC organ transplantation and myocardial infarction. They can also be used
 CC for treating autoimmune related diseases and IL-1 dependent cancers.
 SQ Sequence 159 AA;

Query Match 71.4%; Score 45; DB 35; Length 159;
 Best Local Similarity 42.9%; Pred. No. 1.76e+02;
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 143 mpdegvm 149
 :|::|:
 QY 598 LPEDGIM 604

RESULT 14
 ID W37787 standard; Protein; 159 AA.
 AC W37787;
 DE 28-JUL-1998 (first entry)
 DE Recombinant interleukin-1 receptor antagonist amino acid sequence.
 KW Interleukin-1; IL-1; recombinant; receptor antagonist; cell adhesion;
 KW graft-versus-host disease; autoimmune disease; bone marrow transplant;
 KW leukaemia treatment.
 OS Homo sapiens.
 PN US5747444-A.
 PD 05-MAY-1998.
 PF 09-OCT-1990; 594827.
 PR 08-NOV-1991; US-791474.
 PR 09-OCT-1990; US-594827.
 PR 18-APR-1995; US-425232.
 PR 02-JUN-1995; US-459811.
 PA (CHIR) CHIRON CORP.
 PI Haskill JS, Ralph P;
 DR WPI; 98-285792/25.
 DR N-PSDB; V19029.
 PT Treating graft-versus-host diseases - comprises administering
 PT recombinant interleukin-1 receptor antagonists
 PS Claim 1; Fig 1; 14pp; English.
 CC This is the amino acid sequence of the recombinant intracellular
 CC interleukin-1 (IL-1) receptor antagonist, which prevents binding of
 CC IL-1 to its receptor. In the method of the invention it is used in
 CC the treatment of graft-versus-host disease. It is also used to
 CC prevent undesirable cell adhesion of IL-1 receptor positive cells to
 CC endothelial cells. This is especially useful in autoimmune diseases
 CC or states e.g. graft-versus-host rejection in bone marrow transplants
 CC (in leukaemia treatment).

SQ Sequence 159 AA;

Query Match 71.4%; Score 45; DB 30; Length 159;
Best Local Similarity 42.9%; Pred. No. 1.76e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 143 mpdegvm 149
:|:|:|
QY 598 LPEDGIM 604

RESULT 15

ID P93616 standard; Protein; 165 AA.
AC P93616;
DT 14-MAR-1992 (first entry)
DE Sequence encoded by bps 61-600 of interleukin-1 inhibitor
DE (IL-1) gene.
KW Interleukin-1 inhibitor; inflammation therapy;
OS Immunosuppressive agent; inflamed joint.
OS Homo sapiens.
PN EP-343684-A.
PD 29-NOV-1989.
PF 26-MAY-1989; 109540.
PR 27-MAY-1988; US-199915.
PR 31-AUG-1988; US-238171.
PR 23-AUG-1988; US-238713.
PR 23-SEP-1988; US-248521.
PR 03-NOV-1988; US-266531.
PA (SYNE-) SYNERGEN INC.
PI Hannum CH, Eisenberg SP, Thompson RC, Arend WP, Joslin FG,
PI Sommer A;
DR WPI; 89-349765/48.
DR N-PSDB; N92441.
PT Purified interleukin-1 inhibitor - used as an immuno-suppressing
PT agent or to prevent tissue destruction at sites of inflammation
PS Claim 18; Page 28; 53pp; English.
CC The inventors specifically claim recombinant DNA molecule GT10-IL1i-
CC 2A; and the interleukin-1 inhibitors IL-1i-X, IL-1i-alpha and IL-1i-
CC beta. The preferred component is cDNA or a genomic polynucleotide
CC sequence. It includes bases 99-557 of GT10-IL1i-2A (see N92441-
CC N92443). Also claimed is a purified interleukin-1 inhibitor (IL-1i),
CC which is active against 1 or more than 1 of IL-1alpha and IL-1beta.
SQ Sequence 165 AA;

Query Match 71.4%; Score 45; DB 4; Length 165;
Best Local Similarity 42.9%; Pred. No. 1.76e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 149 mpdegvm 155
:|:|:|
QY 598 LPEDGIM 604

Search completed: Thu Jul 8 19:10:01 1999
Job time : 18 secs.

WQSRH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:09:11 1999; MasPar time 5.21 Seconds
Tabular output not generated. 69.272 Million cell updates/sec

Title: >US-09-041-236-2
Description: (598-606) from US09041236.pep (30 of 45)
Perfect Score: 63
Sequence: 1 LPEDGIMAE 9

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.514; Variance 25.783; scale 0.873

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	48	76.2	182	2 A64007	hypothetical protein	5.68e+00
2	48	76.2	2763	1 WZBE22	gene 22 protein - hum	5.68e+00
3	47	74.6	361	2 PN0100	riboflavin-specific d	9.32e+00
4	47	74.6	499	2 S18965	heat shock protein 82	9.32e+00
5	47	74.6	705	2 A43508	heat shock protein 83	9.32e+00
6	47	74.6	715	2 A48426	heat shock protein HS	9.32e+00
7	46	73.0	314	1 PAFYF	phosphoprotein phosph	1.52e+01
8	46	73.0	323	2 S47741	hypothetical transcri	1.52e+01
9	46	73.0	605	2 A48665	methylnalonyl-CoA car	1.52e+01
10	45	71.4	177	2 A30368	interleukin-1 recepto	2.45e+01
11	45	71.4	180	2 A39386	interleukin-1 recepto	2.45e+01
12	45	71.4	200	2 H71222	hypothetical protein	2.45e+01
13	45	71.4	287	2 G70974	probable cmaA1 protei	2.45e+01
14	45	71.4	293	2 B64339	spermidine synthase (2.45e+01
15	45	71.4	556	2 H69279	adenine deaminase (ad	2.45e+01
16	45	71.4	604	2 S56027	hypothetical protein	2.45e+01
17	44	69.8	242	2 G69128	conserved hypothetical	3.93e+01
18	44	69.8	255	2 S50326	phosphoglycerate muta	3.93e+01
19	44	69.8	443	2 E65008	hypothetical protein	3.93e+01
20	44	69.8	533	2 A43489	RNA-directed RNA poly	3.93e+01
21	44	69.8	556	2 B64482	adenine deaminase (EC	3.93e+01
22	44	69.8	687	2 S51398	hypothetical protein	3.93e+01
23	44	69.8	1932	2 S53409	probable membrane pro	3.93e+01

24	43	68.3	104	2 B69087	conserved hypothetical	6.25e+01
25	43	68.3	180	2 H71043	probable 30S ribosoma	6.25e+01
26	43	68.3	248	2 D29525	methyl coenzyme M red	6.25e+01
27	43	68.3	252	2 S66289	alpha 1 antichymotryp	6.25e+01
28	43	68.3	325	2 A40963	phospholipase A2-acti	6.25e+01
29	43	68.3	339	1 F69852	L-iditol 2-dehydrogen	6.25e+01
30	43	68.3	383	2 JC5775	aspartate transaminas	6.25e+01
31	43	68.3	385	2 JC4537	aspartate transaminas	6.25e+01
32	43	68.3	402	2 I49471	alpha-1 proteinase in	6.25e+01
33	43	68.3	410	2 I50494	serine proteinase inh	6.25e+01
34	43	68.3	413	2 I49452	alpha-1-antitrypsin p	6.25e+01
35	43	68.3	458	2 E65068	hypothetical protein	6.25e+01
36	43	68.3	461	2 A30364	alpha-L-fucosidase (E	6.25e+01
37	43	68.3	578	2 B64012	hypothetical protein	6.25e+01
38	43	68.3	582	2 B70389	translation elongatio	6.25e+01
39	43	68.3	647	2 JC4239	phospholipase A2-acti	6.25e+01
40	43	68.3	675	2 B56262	exonuclease ABC chai	6.25e+01
41	43	68.3	695	2 S28533	tma protein - Lactoco	6.25e+01
42	43	68.3	696	2 S71889	immune factor 1, type	6.25e+01
43	43	68.3	834	2 S54563	hypothetical protein	6.25e+01
44	42	66.7	368	2 S63363	phosphoprotein phosph	9.84e+01
45	42	66.7	577	2 E64328	glycine--tRNA ligase	9.84e+01

ALIGNMENTS

RESULT 1
ENTRY #type complete
TITLE hypothetical protein HI0391 - Haemophilus influenzae (strain Rd KW20)
ORGANISM #formal_name Haemophilus influenzae
DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 10-Oct-1997
ACCESSIONS A64007
REFERENCE A64007
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Keriavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.B.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal Science (1995) 269:496-512
#title Whole-genome random sequencing and assembly of Haemophilus influenzae Rd
#cross-references MDDB:95350630
#accession A64007
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-182 #label TIGR
#cross-references GB:U32722; GB:L42023; NID:gl573348; PID:gl573362; TIGR:HI0391
GENETICS
#start_codon GTG
SUMMARY #length 182 #molecular-weight 21053 #checksum 4179
Query Match 76.2%; Score 48; DB 2; Length 182;
Best Local Similarity 66.7%; Pred. No. 5.68e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 134 LPKDSVMAE 142
Qy 598 LPEDGIMAE 606
RESULT 2
ENTRY #type complete
TITLE gene 22 protein - human herpesvirus 3
ORGANISM #formal_name human herpesvirus 3, varicella-zoster virus
DATE 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change

“


```
ACCESSIONS      A45508
REFERENCE
#authors      Conner, T.W.; LaFayette, P.R.; Nagao, R.T.; Key, J.L.
#journal      Plant Physiol. (1990) 94:1689-1695
#title        Sequence and expression of a HSP83 from Arabidopsis thaliana.
#accession    A45508
##status      preliminary
##molecule_type mRNA
##residues    1-705 #label CON
##cross-references GB:M62984; NID:g166769; PID:g166770
##note        the authors translated the codon AAA for residue 249 as
              Cys, and GGT for residue 425 as Glu
CLASSIFICATION #superfamily heat shock protein 90
SUMMARY        #length 705 #molecular-weight 81073 #checksum 283

Query Match    74.6%; Score 47; DB 2; Length 705;
Best Local Similarity 62.5%; Pred. No. 9.32e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 616 PDNGIMEE 623
|::||| |
QY 599 PEDGIMAE 606

RESULT 6
ENTRY
TITLE      heat shock protein HSP82 - maize
ORGANISM   19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change
DATE       20-Mar-1998
ACCESSIONS A48426
REFERENCE
#authors   Marrs, K.A.; Casey, E.S.; Capitant, S.A.; Bouchard, R.A.;
           Dietrich, P.S.; Mettler, I.J.; Sinibaldi, R.M.
#journal   Dev. Genet. (1993) 14:27-41
#title     Characterization of two maize HSP90 heat shock protein genes:
           expression during heat shock, embryogenesis, and pollen
           development.
#cross-references MUID:93245392
#accession  A48426
##status    preliminary
##molecule_type nucleic acid
##residues  1-715 #label MAR
##cross-references GB:S59780; NID:g300082; PID:g300083
##note      sequence inconsistent with the nucleotide translation
           sequence extracted from NCBI backbone (NCBIN:130885,
           NCBI:P:130886)
CLASSIFICATION #superfamily heat shock protein 90
SUMMARY        #length 715 #molecular-weight 81919 #checksum 1254

Query Match    74.6%; Score 47; DB 2; Length 715;
Best Local Similarity 62.5%; Pred. No. 9.32e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 626 PDNGIMEE 633
|::||| |
QY 599 PEDGIMAE 606

RESULT 7
ENTRY
TITLE      phosphoprotein phosphatase (EC 3.1.3.16) Y - fruit fly
           (Drosophila melanogaster)
ORGANISM   31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
DATE       29-May-1998
ACCESSIONS S03963; S11061
REFERENCE
#authors   Dombradi, V.; Axton, J.M.; Glover, D.M.; Cohen, P.T.W.
#journal   FEBS Lett. (1989) 247:391-395
#title     Molecular cloning and chromosomal localization of a novel
           Drosophila protein phosphatase.
```

```
#cross-references MUID:89232161
#accession    S03963
##molecule_type mRNA
##residues    1-314 #label DOM
##cross-references EMBL:Y07510; NID:g8371; PID:g8372
REFERENCE     S11059
#authors      Cohen, P.T.W.; Breviis, N.D.; Hughes, V.; Mann, D.J.
#journal      FEBS Lett. (1990) 268:355-359
#title        Protein serine/threonine phosphatases; an expanding family.
#cross-references MUID:90346193
#accession    S11061
##status      nucleic acid sequence not shown
##molecule_type mRNA
##residues    1-314 #label COH
GENETICS
#gene         FlyBase:PyY-55A
#map_position 2R 55A1-3
##cross-references FlyBase:FBgn0003140
CLASSIFICATION #superfamily phosphoprotein phosphatase; phosphoesterase core
               homology; phosphoprotein phosphatase homology
               iron; phosphoric monoester hydrolase;
               serine/threonine-specific phosphatase; zinc
KEYWORDS
FEATURE
26-285        #domain phosphoprotein phosphatase homology #label PPP\
60,62,88       #binding_site iron (Asp, His, Asp) #status predicted\
88,120,169,244 #binding_site zinc (Asp, Asn, His, His) #status
               predicted\
91,121,268     #active_site Asp, His, Tyr #status predicted\
92,217         #binding_site substrate phosphate (Arg) #status
               predicted
SUMMARY        #length 314 #molecular-weight 36022 #checksum 1802

Query Match    73.0%; Score 46; DB 1; Length 314;
Best Local Similarity 44.4%; Pred. No. 1.52e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 191 IPDEGIMCD 199
|::||| |
QY 598 LPEDGIMAE 606

RESULT 8
ENTRY
TITLE      hypothetical transcription regulator tref-kdgK intergenic
           region - Escherichia coli
ALTERNATE_NAMES hypothetical protein o323
ORGANISM      #formal_name Escherichia coli
DATE          27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change
               02-Jul-1998
ACCESSIONS    S47741; D65150
REFERENCE     S47666
#authors      Plunkett, G.
#submission   submitted to the EMBL Data Library, March 1994
#accession    S47741
##status      preliminary
##molecule_type DNA
##residues    1-323 #label PLU
##cross-references EMBL:U00039; NID:g466582; PID:g46658
REFERENCE     A64720
#authors      Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
               Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
               Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
               Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
               Y.
#journal      Science (1997) 277:1453-1462
#title        The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession    D65150
##status      preliminary; nucleic acid sequence not shown;
               translation not shown
##molecule_type DNA
##residues    1-323 #label BLAT
##cross-references GB:AE000428; GB:U00096; NID:g1789931; PID:g1789938;
```

```

# #experimental_source strain K-12, substrain MG1655
GENETICS
#gene
SUMMARY
#length 323 #molecular-weight 36093 #checksum 3830
Query Match 73.0%; Score 46; DB 2; Length 323;
Best Local Similarity 75.0%; Pred. No. 1.52e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 172 LPDGVIA 179
QY 598 LPEDGIMA 605

RESULT 9
ENTRY
TITLE
methylenalonyl-CoA carboxyltransferase (EC 2.1.3.1) 12S chain
- Propionibacterium freudenreichii subsp. shermanii
ORGANISM
#formal_name Propionibacterium freudenreichii subsp.
#Shermanii
DATE
07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
17-Mar-1999
ACCESSIONS
A48665
REFERENCE
A48665
#authors
Thornton, C.G.; Kumar, G.K.; Haase, F.C.; Phillips, N.F.B.;
Woo, S.B.; Park, V.M.; Magner, W.J.; Shenoy, B.C.; Wood,
H.G.; Samols, D.
#journal
J. Bacteriol. (1993) 175:5301-5308
#title
Primary structure of the monomer of the 12S subunit of
transcarboxylase as deduced from DNA and characterization
of the product expressed in Escherichia coli.
#cross-references MUID:93374821
#accession
A48665
#status
Preliminary
#molecule_type DNA
#residues 1-605 #label THO
#cross-references GB:L04196; NID:gi50932; PID:gi50933
CLASSIFICATION
#superfamily propionyl-CoA carboxylase beta chain
KEYWORDS
transferase
SUMMARY
#length 605 #molecular-weight 65699 #checksum 5947

Query Match 73.0%; Score 46; DB 2; Length 605;
Best Local Similarity 55.6%; Pred. No. 1.52e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 519 LSSEEIMAD 527
QY 598 LPEDGIMAE 606

RESULT 10
ENTRY
TITLE
Interleukin-1 receptor antagonist secreted form precursor -
human
ORGANISM
#formal_name Homo sapiens #common_name man
DATE
07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change
13-Sep-1998
ACCESSIONS
A40956; A30368; S08160; S08159; A37822
REFERENCE
A40956
#authors
Eisenberg, S.P.; Brewer, M.T.; Verderber, E.; Heimdal, P.;
Brandhuber, B.J.; Thompson, R.C.
#journal
Proc. Natl. Acad. Sci. U.S.A. (1991) 88:5232-5236
#title
Interleukin 1 receptor antagonist is a member of the
interleukin 1 gene family: evolution of a cytokine control
mechanism.
#cross-references MUID:91271363
#accession
A40956
#molecule_type DNA
#residues 1-177 #label EIS
#cross-references GB:M63099; NID:gi186385; PID:gi186386
REFERENCE
137894
#authors
Lennard, A.; Gorman, P.; Carrier, M.; Griffiths, S.; Sootney,

```

```

H.; Sheer, D.; Solari, R.
Cytokine (1992) 4:83-89
Cloning and chromosome mapping of the human interleukin-1
receptor antagonist gene.
#cross-references MUID:92338323
#accession
I37894
#status
translated from GB/EMBL/DBBJ
#molecule_type DNA
#residues 1-177 #label LEN
#cross-references EMBL:X64532; NID:933798; PID:933799
A30368
REFERENCE
#authors
Carter, D.B.; Deibel Jr., M.R.; Dunn, C.J.; Tomich, C.S.C.;
Laborde, A.L.; Slightom, J.L.; Berger, A.E.; Bienkowski,
M.J.; Sun, F.F.; McEwan, R.N.; Harris, P.K.W.; Yen, A.W.;
Waszak, G.A.; Chosay, J.G.; Siew, L.C.; Hardee, M.M.;
Zurcher-Neely, H.A.; Reardon, I.M.; Heinrichson, R.L.;
Truesdell, S.E.; Shelly, J.A.; Eessalu, T.E.; Taylor, B.M.;
Tracey, D.E.
#journal
Nature (1990) 344:633-638
#title
Purification, cloning, expression and biological
characterization of an interleukin-1 receptor antagonist
protein.
#cross-references MUID:90220867
#accession
A30368
#molecule_type mRNA
#residues 1-177 #label CAR
#cross-references GB:X53296; NID:932578; PID:932579
#note
parts of this sequence, including the amino end of the
mature protein, were confirmed by peptide sequencing
REFERENCE
S08160
#authors
Eisenberg, S.P.; Evans, R.J.; Arend, W.P.; Verderber, E.;
Brewer, M.T.; Hannum, C.H.; Thompson, R.C.
#journal
Nature (1990) 343:341-346
#title
Primary structure and functional expression from
complementary DNA of a human interleukin-1 receptor
antagonist.
#cross-references MUID:90136921
#accession
S08160
#status
not compared with conceptual translation
#molecule_type mRNA
#residues 1-177 #label E12
#cross-references GB:X52015; NID:932576; PID:932577
REFERENCE
S08159
#authors
Hannum, C.H.; Wilcox, C.J.; Arend, W.P.; Joslin, F.G.;
Dripps, D.J.; Heimdal, P.L.; Armes, L.G.; Sommer, A.;
Eisenberg, S.P.; Thompson, R.C.
#journal
Nature (1990) 343:336-340
#title
Interleukin-1 receptor antagonist activity of a human
interleukin-1 inhibitor.
#cross-references MUID:90136920
#accession
S08159
#molecule_type protein
#residues 26-75;97-108;110-116;120-131;163-176 #label HAN
REFERENCE
A37822
#authors
Bienkowski, M.J.; Eessalu, T.E.; Berger, A.E.; Truesdell,
S.E.; Shelly, J.A.; Laborde, A.L.; Zurcher-Neely, H.A.;
Reardon, I.M.; Heinrichson, R.L.; Chosay, J.G.; Tracey, D.E.
#journal
J. Biol. Chem. (1990) 265:14503-14511
#title
Purification and characterization of interleukin 1 receptor
level antagonist proteins from THP-1 cells.
#cross-references MUID:90354444
#accession
A37822
#molecule_type protein
#residues 26-52;70-77;122-127;170-175 #label BIE
#experimental_source culture medium, PMA-stimulated THP-1 cells
GENETICS
#gene
GDB:IL1RN
#cross-references GDB:125897; OMIM:147679
#map_position 2q14.2-2q14.2
#introns 39/2; 69/1; 106/3
CLASSIFICATION
#superfamily interleukin-1
alternative splicing; cytokine receptor; extracellular
protein; glycoprotein

```

```

FEATURE
1-25      #domain signal sequence #status predicted #label SIG\
26-177    #product interleukin-1 receptor antagonist #status
          experimental #label MA1\
109       #binding_site carbohydrate (Asn) (covalent) #status
          experimental
SUMMARY   #length 177 #molecular-weight 20055 #checksum 193

Query Match      71.4%; Score 45; DB 2; Length 177;
Best Local Similarity 42.9%; Pred. No. 2.45e+01;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 161 MPDEGVM 167
QY 598 LPEDGIM 604
      :|::|:|

RESULT 11
ENTRY   #type complete
TITLE   interleukin-1 receptor antagonist, long intracellular splice
        form - human
CONTAINS
interleukin-1 receptor antagonist, short intracellular splice
form
#formal_name Homo sapiens #common name man
#formal_name Homo sapiens #sequence_revision 11-Apr-1997 #text_change
24-Sep-1998
ACCESSIONS  I37893; A39386
REFERENCE    I37893
#authors     Muzio, M.; Polentarutti, N.; Sironi, M.; Poli, G.; De Gioia,
L.; Introna, M.; Mantovani, A.; Colotta, F.
#journal     J. Exp. Med. (1995) 182:623-628
#title       Cloning and characterization of a new isoform of the
interleukin 1 receptor antagonist.
#cross-references MUID:95355865
#accession     I37893
#status        translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues      1-180 #label RES
#cross-references EMBL:X84348; NID:g1008970; PID:g1008971
REFERENCE     A39386
#authors      Haskill, S.; Martin, G.; Van Le, L.; Morris, J.; Peace, A.;
Biglier, C.F.; Jaffe, G.J.; Hammerberg, C.; Sporn, S.A.;
Fong, S.; Arend, W.P.; Ralph, P.
#journal     Proc. Natl. Acad. Sci. U.S.A. (1991) 88:3681-3685
#title       cDNA cloning of an intracellular form of the human
interleukin 1 receptor antagonist associated with
epithelium.
#cross-references MUID:91219436
#accession     A39386
#molecule_type mRNA
#residues      1-3,25-180 #label HAS
#cross-references GB:M55646; NID:g186291; PID:g186292
GENETICS
#gene         GDB:IL1RN
#map_position 2q14.2-2q14.2
CLASSIFICATION #superfamily interleukin-1
KEYWORDS       alternative splicing; cytokine receptor
FEATURE
1-180
#product interleukin-1 receptor antagonist, long
intracellular splice form #status predicted #label
MA1\
#product interleukin-1 receptor antagonist, short
intracellular splice form #status predicted #label MA2
SUMMARY   #length 180 #molecular-weight 19897 #checksum 956

Query Match      71.4%; Score 45; DB 2; Length 180;
Best Local Similarity 42.9%; Pred. No. 2.45e+01;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 164 MPDEGVM 170
QY 598 LPEDGIM 604
      :|::|:|

FEATURE
1-25      #domain signal sequence #status predicted #label SIG\
26-177    #product interleukin-1 receptor antagonist #status
          experimental #label MA1\
109       #binding_site carbohydrate (Asn) (covalent) #status
          experimental
SUMMARY   #length 177 #molecular-weight 20055 #checksum 193

Query Match      71.4%; Score 45; DB 2; Length 177;
Best Local Similarity 42.9%; Pred. No. 2.45e+01;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 161 MPDEGVM 167
QY 598 LPEDGIM 604
      :|::|:|

RESULT 12
ENTRY   #type complete
TITLE   hypothetical protein PH0043 - Pyrococcus horikoshii
        #formal_name Pyrococcus horikoshii
        14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
        05-Dec-1998
ACCESSIONS  A71000
REFERENCE    A71000
#authors     Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#journal     DNA Res. (1998) 5:55-76
#title       Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references MUID:98344137
#accession     H71222
#status        preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues      1-200 #label KAW
#cross-references GB:AP000001; NID:g2326128; PID:d1030054; PID:g3256428
#experimental_source strain OT3
#note          this accession replaces an interim accession for a
sequence replaced by GenBank
GENETICS
#gene         PH0043
CLASSIFICATION #superfamily Methanobacterium thermoautotrophicum conserved
hypothetical protein MTH1318
SUMMARY   #length 200 #molecular-weight 22277 #checksum 380

Query Match      71.4%; Score 45; DB 2; Length 200;
Best Local Similarity 66.7%; Pred. No. 2.45e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 30 LPGDGVIAE 38
QY 598 LPEDGIMAE 606
      |||::|:|

RESULT 13
ENTRY   #type complete
TITLE   Probable cmaAI protein - Mycobacterium tuberculosis (strain
        H37RV)
        #formal_name Mycobacterium tuberculosis
        17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
        17-Jul-1998
ACCESSIONS  G70974
REFERENCE    A70500
#authors     Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Basham, D.; Brown, D.;
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, L.; Jagers, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal     Nature (1998) 393:537-544
#title       Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession     G70974
#status        preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA

```

```
##residues      1-287 ##label COL
##cross-references GB:AL009198; GB:AL123456; NID:g3242262; PID:el102310;
PID:g2661671
##experimental_source strain H37Rv
```

GENETICS

```
#gene
```

```
SUMMARY
```

```
  #length 287  #molecular-weight 32461  #checksum 7948
```

```
Query Match      71.4%; Score 45; DB 2; Length 287;
Best Local Similarity 71.4%; Pred. No. 2.45e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Db 158 LPADGVM 164
```

```
  || |||
```

```
QY 598 LPEDGIM 604
```

RESULT 14

```
ENTRY
```

```
TITLE
```

```
ORGANISM
```

```
DATE
```

```
ACCESSIONS
```

```
REFERENCE
```

```
#authors
```

```
B64339
```

```
A64300
```

```
Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
```

```
R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
```

```
R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
```

```
Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
```

```
Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
```

```
Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann,
```

```
J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
```

```
J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
```

```
K.M.; Hurst, C.M.; Kaine, B.P.; Borodovsky, M.; Klenk,
```

```
H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
```

```
Science (1996) 273:1038-1073
```

```
Complete genome sequence of the methanogenic archaeon,
```

```
Methanococcus jannaschii.
```

```
#cross-references MUID:96337999
```

```
#accession
```

```
B64339
```

```
##status
```

```
preliminary; nucleic acid sequence not shown;
```

```
translation not shown
```

```
##molecule_type DNA
```

```
##residues 1-293 ##label BUL
```

```
##cross-references GB:U67486; GB:L77117; NID:gl591031; PID:gl591033;
```

```
TIGR:MU0313; PID:gl510404
```

```
GENETICS
```

```
#map_position REV293438-292557
```

```
#start_codon GTG
```

```
CLASSIFICATION
```

```
KEYWORDS
```

```
SUMMARY
```

```
  #length 293  #molecular-weight 33899  #checksum 9220
```

```
Query Match      71.4%; Score 45; DB 2; Length 293;
```

```
Best Local Similarity 55.6%; Pred. No. 2.45e+01;
```

```
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Db 187 LNDDGIMVQ 195
```

```
  | : ||||
```

```
QY 598 LPEDGIMAE 606
```

RESULT 15

```
ENTRY
```

```
TITLE
```

```
ORGANISM
```

```
DATE
```

```
ACCESSIONS
```

```
REFERENCE
```

```
#authors
```

```
H69279
```

```
A69250
```

```
Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
```

D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artiaich, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.

Nature (1997) 390:364-370

The complete genome sequence of the hyperthermophilic,

sulfate-reducing archaeon *Archaeoglobus fulgidus*.

#cross-references MUID:98049343

#accession

H69279

##status

preliminary; nucleic acid sequence not shown;

translation not shown

##molecule_type DNA

##residues 1-556 ##label KLE

##cross-references GB:AE001089; GB:AE000782; NID:g2689412; PID:g2650401;

TIGR:AF0240

CLASSIFICATION

#superfamily adenine deaminase adeC

#length 556 #molecular-weight 60764 #checksum 978

SUMMARY

Query Match 71.4%; Score 45; DB 2; Length 556;

Best Local Similarity 55.6%; Pred. No. 2.45e+01;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 485 LPDIAGMSD 493

|| |||

QY 598 LPEDGIMAE 606

Search completed: Thu Jul 8 19:09:23 1999

Job time : 12 secs.

MORFAL (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mperch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:08:13 1999; MasPar time 3.39 Seconds
75.127 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (598-606) from US09041236.pep (30 of 45)
Perfect Score: 63
Sequence: 1 LPEDGIMAE 9

Scoring table: PAM 150
Gap 11

Searched: 77977 segs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 23.146; Variance 22.898; scale 1.011

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	48	76.2	182	1	Y391_HAEIN HYPOTHETICAL PROTEIN H	1.76e+00
2	48	76.2	2763	1	TEGU_VZVD LARGE TEGUMENT PROTEIN	1.76e+00
3	47	74.6	256	1	SUBB_THEME EXTRAGENIC SUPPRESSOR	3.08e+00
4	47	74.6	361	1	RIBG_BACSU RIBOFLAVIN-SPECIFIC DE	3.08e+00
5	47	74.6	499	1	HS82_TOBAC HEAT SHOCK PROTEIN 82	3.08e+00
6	47	74.6	700	1	HS81_ARATH HEAT SHOCK PROTEIN 81	3.08e+00
7	47	74.6	703	1	HS83_PHANI HEAT SHOCK PROTEIN 83	3.08e+00
8	47	74.6	715	1	HS82_MAIZE HEAT SHOCK PROTEIN 82	3.08e+00
9	47	74.6	743	1	TFE3_HUMAN TRANSCRIPTION FACTOR E	3.08e+00
10	45	73.0	299	1	YHJC_ECOLI HYPOTHETICAL TRANSCRIP	5.31e+00
11	46	73.0	314	1	PPY_DROME SERINE/THREONINE PROTE	5.31e+00
12	45	71.4	177	1	IL1X_HUMAN INTERLEUKIN-1 RECEPTOR	9.09e+00
13	45	71.4	287	1	CFAL_MYCTU CYCLOPROPANE-FATTY-ACY	9.09e+00
14	45	71.4	293	1	SPEE_METJA PUTATIVE SPERMIDINE SY	9.09e+00
15	45	71.4	371	1	RIBG_BACAM RIBOFLAVIN-SPECIFIC DE	9.09e+00
16	45	71.4	556	1	ADEC_ARCFU PROBABLE ADENINE DEAM	9.09e+00
17	45	71.4	604	1	IYJ3_YEAST HYPOTHETICAL 68.8 KD P	9.09e+00
18	44	69.8	239	1	YTEA_BACSU HYPOTHETICAL 27.7 KD P	1.54e+01
19	44	69.8	359	1	GLNA_SCHPO GLUTAMINE SYNTHETASE (1.54e+01
20	44	69.8	453	1	UFQG_GENTR FLAVONOL 3-O-GLUCOSYLT	1.54e+01
21	44	69.8	533	1	RRPB_CVPFS RNA-DIRECTED RNA POLYM	1.54e+01
22	44	69.8	556	1	ADEC_METJA PROBABLE ADENINE DEAM	1.54e+01
23	43	68.3	247	1	MCRG_METBA METHYL-COENZYME M REDU	2.58e+01

24	43	68.3	413	1	AL12_MOUSE ALPHA-1 ANTITRYPSIN 1-	2.58e+01
25	43	68.3	458	1	YGER_ECOLI HYPOTHETICAL 52.8 KD P	2.58e+01
26	43	68.3	461	1	FUCO_DICDI ALPHA-L-FUCOSIDASE PRE	2.58e+01
27	43	68.3	578	1	YTFM_HAEIN HYPOTHETICAL PROTEIN H	2.58e+01
28	43	68.3	646	1	PHOSPHOLIPASE A-2-ACTI	2.58e+01
29	43	68.3	647	1	PLAP_RAT PHOSPHOLIPASE A-2-ACTI	2.58e+01
30	43	68.3	675	1	UVRE_NEIME EXCINUCLEASE ABC SUBUN	2.58e+01
31	43	68.3	675	1	UVRE_NEIGO EXCINUCLEASE ABC SUBUN	2.58e+01
32	43	68.3	695	1	FTSH_LACLA CELL DIVISION PROTEIN	2.58e+01
33	43	68.3	834	1	YM22_YEAST HYPOTHETICAL 94.3 KD T	2.58e+01
34	43	68.3	1918	1	CC4H_HUMAN CDC4-LIKE PROTEIN (FRA	2.58e+01
35	42	66.7	86	1	Y0BD_MYCLE HYPOTHETICAL 9.4 KD PR	4.28e+01
36	42	66.7	212	1	AL16_MOUSE ALPHA-1 ANTITRYPSIN 1-	4.28e+01
37	42	66.7	252	1	YQIH_ECOLI HYPOTHETICAL FIMBRIAL	4.28e+01
38	42	66.7	265	1	Y176_HUMAN HYPOTHETICAL PROTEIN K	4.28e+01
39	42	66.7	271	1	YGID_ECOLI HYPOTHETICAL 29.9 KD P	4.28e+01
40	42	66.7	412	1	AL12_MUSCR ALPHA-1 ANTITRYPSINASE	4.28e+01
41	42	66.7	413	1	AL13_MOUSE ALPHA-1 ANTITRYPSIN 1-	4.28e+01
42	42	66.7	577	1	SYG_METJA PROBABLE GLYCYL-TRNA S	4.28e+01
43	42	66.7	699	1	HS80_LICES HEAT SHOCK COGNATE PRO	4.28e+01
44	42	66.7	1220	1	ATCP_RABIT CALCITUM-TRANSPORTING A	4.28e+01
45	42	66.7	3685	1	DMD_HUMAN DYSTROPHIN	4.28e+01

ALIGNMENTS

RESULT	ID	Y391_HAEIN	STANDARD;	PRT;	182 AA.
AC	P43992				
DT	01-NOV-1995	(REL. 32, CREATED)			
DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)			
DT	01-NOV-1995	(REL. 32, LAST ANNOTATION UPDATE)			
DE	HYPOTHETICAL PROTEIN HI0391.				
GN	HI0391.				
OS	HAEMOPHILUS INFLUENZAE.				
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;				
OC	HAEMOPHILUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-RD / KW20;				
RX	MEDLINE: 95350630.				
RA	FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,				
RA	KERLAVAGE A.R., BULT C.J., TOMB J.F., DOUGHERTY B.A., MERRICK J.M.,				
RA	MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,				
RA	SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,				
RA	WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,				
RA	UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,				
RA	FINE L.D., FRITZMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,				
RA	GNERM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,				
RA	VENTER J.C.;				
RT	"Whole-genome random sequencing and assembly of Haemophilus				
RT	influenzae Rd.,"				
RL	SCIENCE 269:496-512(1995).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U32722; G1573362; -.				
DR	TIGR; HI0391; -.				
KW	HYPOTHETICAL PROTEIN				
SQ	SEQUENCE 182 AA; 21053 MW; 32475BEE CRC32;				

Query Match 76.2%; Score 48; DB 1; Length 182;
Best Local Similarity 66.7%; Pred. No. 1.76e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 134 LPEDSVMAE 142
|||::|||

```
Oy 598 LPEDGIMAE 606

RESULT 2
ID TEGU_VZVD STANDARD; PRT: 2763 AA.
AC P09278;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE LARGE TEGUMENT PROTEIN.
GN 22
OS VARICELLA-ZOSTER VIRUS (STRAIN DUMAS) (VZV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86306657.
RA DAVISON A.J., SCOTT J.E.;
RT "The complete DNA sequence of varicella-zoster virus.";
RL J. GEN. VIROL. 67:1759-1816(1986).
CC -!- FUNCTION: TEGUMENT PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EH-1 24, EBV BFL1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04370; G60011; -
DR PIR; D27343; WZBE22.
DR PIR; D27343; WZBE22.
SQ SEQUENCE 2763 AA; 306339 MW; E550B4D8 CRC32;

Query Match 76.2%; Score 48; DB 1; Length 2763;
Best Local Similarity 55.6%; Pred. No. 1.76e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1907 LPEDGIVEE 1915
||:|:|:|
Oy 598 LPEDGIMAE 606

RESULT 3
ID SUBH_THEME STANDARD; PRT: 256 AA.
AC O33832;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE EXTRAGENIC SUPPRESSOR PROTEIN SUBH HOMOLOG.
GN SUBH.
OS THERMOTOGA MARITIMA.
OC BACTERIA; THERMOTOGALES; THERMOTOGA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN MSB8 / DSM 3109;
RL LIEBL W., BREM D., GOTSCHLICH A.;
RA SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ001073; E33354; -
DR PROSITE; PS00629; IMP_1; 1.
DR PROSITE; PS00630; IMP_2; 1.

DR PFAM; PF00459; inositol_P; 1.
SQ SEQUENCE 256 AA; 28647 MW; 2F9A8E44 CRC32;

Query Match 74.6%; Score 47; DB 1; Length 256;
Best Local Similarity 55.6%; Pred. No. 3.08e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 57 FPDENIMAE 65
||:|:|:|
Oy 598 LPEDGIMAE 606

RESULT 4
ID RIBG_BACSU STANDARD; PRT: 361 AA.
AC P17618;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOFLAVIN-SPECIFIC DEAMINASE (EC 3.5.4.-).
GN RIBG.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / MARBURG;
RX MEDLINE; 95020338.
RA SOROKIN A.V., ZUMSTEIN E., AZEVEDO V., EHRLICH S.D., SERROR P.;
RT "The organization of the Bacillus subtilis 188 chromosome region
RT between the spoVA and sera genetic loci, based on sequence data.";
RL MOL. MICROBIOL. 10:385-395(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / SHGW;
RA MIRONOV V.N.;
RL THESIS (1989), USSR ACADEMY OF SCIENCES, RUSSIA.
RN [3]
RP SEQUENCE OF 1-12 FROM N.A.
RX STRAIN=168 / SHGW;
RX MEDLINE; 90271920.
RA MIRONOV V.N., PERUMOV D.A., KRAEV A.S., STEPANOV A.I., SKRYABIN K.G.;
RT "Unusual structure of the regulatory region of the riboflavin
RT biosynthesis operon in Bacillus subtilis.";
RL MOL. BIOL. (MOSK) 24:256-261(1990).
CC -!- FUNCTION: INVOLVED IN RIBOFLAVIN BIOSYNTHESIS. CONVERTS
CC 2,5-DIAMINO-6-(RIBOSYLAMINO)-4(3H)-PYRIMIDINONE 5'-PHOSPHATE
CC INTO 5'-AMINO-6-(RIBOSYLAMINO)-2,4(1H,3H)-PYRIMIDINEDIONE
CC 5'-PHOSPHATE.
CC -!- COFACTOR: ZINC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY. STRONG, TO OTHER RIBOFLAVIN-SPECIFIC DEAMINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L09228; G410125; -
DR EMBL; X51510; G40085; -
DR EMBL; 295116; E1185597; -
DR PIR; A35711; A35711.
DR PIR; S45543; S45543.
DR SUBTILIST; BG10518; RIBG.
DR PROSITE; PS00903; CYT_DCMF_DEAMINASES; 1.
DR PFAM; PF00383; dCMF_cyt_deam; 1.
KW RIBOFLAVIN BIOSYNTHESIS; HYDROLASE; ZINC.
FT METAL 49 49 ZINC (BY SIMILARITY).
FT METAL 74 74 ZINC (BY SIMILARITY).
FT METAL 83 83 ZINC (BY SIMILARITY).
SQ SEQUENCE 361 AA; 39305 MW; A5B0584E CRC32;
```

```
Query Match      74.6%; Score 47; DB 1; Length 361;
Best Local Similarity 62.5%; Pred. No. 3.08e+00;
Matches      5; Conservative      3; Mismatches      0; Indels      0; Gaps      0;

Db 279 LAEGIMS 286
    I::I::I::
QY 598 LPEDGIMA 605

RESULT 5
ID HS82 TOBAC STANDARD; PRT: 499 AA.
AC P36182;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HEAT SHOCK PROTEIN 82 (FRAGMENT).
GN HSP82.
OS NICOTIANA TABACUM (COMMON TOBACCO).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANACEAE; NICOTIANA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. SAMSUN;
RA SEVERIN K., ROTKE R., BEHRENS H., HELLER P., SCHOEFFEL F.;
RL SUBMITTED (NOV-1991) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
    (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X63195; G19880; -.
DR PIR; S18865; S18865.
DR PROSITE; PS00298; HSP90; PARTIAL.
DR PFAM; PF00183; HSP90; 1.
KW CHAPERONE; ATP-BINDING; HEAT SHOCK.
FT NON_TER 1
SQ SEQUENCE 499 AA; 58021 MW; 689FBA81 CRC32;

Query Match      74.6%; Score 47; DB 1; Length 499;
Best Local Similarity 62.5%; Pred. No. 3.08e+00;
Matches      5; Conservative      2; Mismatches      1; Indels      0; Gaps      0;

Db 411 PDNGIMEE 418
    I::I::I::
QY 599 PEDGIMAE 606

RESULT 6
ID HS81 ARATH STANDARD; PRT: 700 AA.
AC P27323; Q03930;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HEAT SHOCK PROTEIN 81-1 (HSP81-1) (HEAT SHOCK PROTEIN 83).
GN HSP81-1 OR HSP83.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA; TISSUE=LEAF;
RA CONNER T.W., LAFAYETTE P.R., NAGAO R.T., KEY J.L.;
```

```
"Sequence and expression of a HSP83 from Arabidopsis thaliana.";
PLANT PHYSIOL. 94:1689-1695(1990).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA; TISSUE=SEED;
RX MEDLINE; 95211380.
RA YABE N., TAKAHASHI T., KOMEDA Y.;
RT "Analysis of tissue-specific expression of Arabidopsis thaliana
HSP90-family gene Hsp81.";
RL PLANT CELL PHYSIOL. 35:1207-1219(1994).
CC -!- FUNCTION: MOLECULAR CHAPERONE. DUE TO ITS ASSOCIATION WITH
CERTAIN PROTEINS SUCH AS HORMONE RECEPTORS AND SOME CLASSES OF
KINASES, IT IS IMPLICATED IN SIGNAL TRANSDUCTION AND DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- TISSUE SPECIFICITY: EXPRESSED CONSTITUTIVELY IN ROOTS ONLY. AFTER
HEAT TREATMENT, EXPRESSED IN MOST TISSUES. LEVELS ALSO INCREASE
AFTER HEAVY METAL TREATMENT.
CC -!- INDUCTION: IN CONTRAST TO OTHER MAJOR HEAT SHOCK PROTEINS, THIS
ONE IS ALSO EXPRESSED AT NORMAL GROWTH TEMPERATURES.
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M62984; G166770; ALT_INIT.
DR EMBL; D00710; G217855; -.
DR PIR; A45508; A45508.
DR PROSITE; PS00298; HSP90; 1.
DR PFAM; PF00183; HSP90; 1.
DR HSP90; P07900; 1VER.
KW CHAPERONE; ATP-BINDING; HEAT SHOCK; MULTIGENE FAMILY.
FT DOMAIN 247 253 POLY-LYS.
FT CONFLICT 72 72 A -> S (IN REF. 2).
FT CONFLICT 175 175 S -> T (IN REF. 2).
FT CONFLICT 210 210 T -> I (IN REF. 2).
FT CONFLICT 281 281 S -> Y (IN REF. 2).
SQ SEQUENCE 700 AA; 80529 MW; 96EC9566 CRC32;

Query Match      74.6%; Score 47; DB 1; Length 700;
Best Local Similarity 62.5%; Pred. No. 3.08e+00;
Matches      5; Conservative      2; Mismatches      1; Indels      0; Gaps      0;

Db 611 PDNGIMEE 618
    I::I::I::
QY 599 PEDGIMAE 606

RESULT 7
ID HS83 PHANI STANDARD; PRT: 703 AA.
AC P51819;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HEAT SHOCK PROTEIN 83.
GN HSP83A.
OS PHARBITIS NIL (VIOLET) (JAPANESE MORNING GLORY).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANACEAE; IPOMOEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COTYLEDON;
RA FELSHEIM R.F., DAS A.;
RT "Structure and expression of a heat shock protein 83 (hsp83) gene of
Pharbitis nil."
RL PLANT PHYSIOL. 100:801-811(1992).
CC -!- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
```

CC (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -|- INDUCTION: BY HEAT SHOCK AND LIGHT. THIS HEAT SHOCK PROTEIN IS
CC INDUCED BY LIGHT ONLY AFTER A PROLONGED DARK PERIOD OR VICE
CC VERSA.
CC -|- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M99431; G169296; -.
CC PROSITE: PS00298; HSP90: 1.
CC PFAM: PF00183; HSP90: 1.
CC HSP: P07900; 1YER.
CC KW CHAPERONE: ATP-BINDING; HEAT SHOCK; MULTIGENE FAMILY.
CC SQ SEQUENCE 703 AA; 80820 MW; 27DE8B48 CRC32;

Query Match 74.6%; Score 47; DB 1; Length 703;
Best Local Similarity 62.5%; Pred. No. 3.08e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 615 PDNGIMEE 622
I:|||||
QY 599 PEDGIMAE 606

RESULT 8
ID HS82-MAIZE STANDARD; PRT; 715 AA.
AC Q08277;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE HEAT SHOCK PROTEIN 82.
GN HSP82.
OS ZEA MAYS (MAIZE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; ZEA.
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=SEEDLING, AND LEAF;
RX MEDLINE: 93245392.
RA MARRS K.A., CASEY E.S., CAPITANT S.A., BOUCHARD R.A.,
RA DIETRICH P.S., METTLER I.J., SINIBALDI R.M.;
RA "Characterization of two maize HSP90 heat shock protein genes:
RA expression during heat shock, embryogenesis, and pollen
RA development.";
RL DEV. GENET. 14:27-41(1993).
CC -|- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
CC (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S59780; E81011; -.
CC MAZEDB: 65833; -.
CC PROSITE: PS00298; HSP90: 1.
CC PFAM: PF00183; HSP90: 1.
CC HSP: P07900; 1YER.
CC KW CHAPERONE: ATP-BINDING; HEAT SHOCK.
CC SQ SEQUENCE 715 AA; 81890 MW; 692A59A5 CRC32;

Query Match 74.6%; Score 47; DB 1; Length 715;
Best Local Similarity 62.5%; Pred. No. 3.08e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 626 PDNGIMEE 633
I:|||||
QY 599 PEDGIMAE 606

RESULT 9
ID TFE3_HUMAN STANDARD; PRT; 743 AA.
AC P19532; Q99964; Q92757; Q92758;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION FACTOR E3.
GN TFE3.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RN SEQUENCE OF 1-219 FROM N.A., AND CHROMOSOMAL TRANSLOCATION.
RX MEDLINE: 97140324.
RA WETERMAN M.A.J., WILBRINK M., GEURTS VAN KESSEL A.;
RT "Fusion of the transcription factor TFE3 gene to a novel gene, PRCC,
RT in t(X;1)(p11;q21)-positive papillary renal cell carcinomas.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:15294-15298(1996).
RN [2]
RP SEQUENCE OF 149-743 FROM N.A., AND CHROMOSOMAL TRANSLOCATION.
RC TISSUE=MONOCYTES;
RX MEDLINE: 97026295.
RA SIDHAR S.K., CLARK J., GILL S., HAMOUDI R., CREW A.J.,
RA WILLIAM R., ROSS M., LINEHAN W.M., BIRDSALL S., SHIPLEY J.,
RA COOPER C.S.;
RT "The t(X;1)(p11;q21.2) translocation in papillary renal cell
RT carcinoma fuses a novel gene PRCC to the TFE3 transcription factor
RT gene.";
RL HUM. MOL. GENET. 5:1333-1338(1996).
RN [3]
RP SEQUENCE OF 208-743 FROM N.A.
RX MEDLINE: 90249724.
RA BECKMANN H., SU L.-K., KADESCH T.;
RT "TFE3: a helix-loop-helix protein that activates transcription
RT through the immunoglobulin enhancer muE3 motif.";
RL GENES DEV. 4:167-179(1990).
CC -|- FUNCTION: POSITIVE-ACTING TRANSCRIPTION FACTOR THAT BINDS TO THE
CC IMMUNOGLOBULIN ENHANCER MUE3 MOTIF. IT BINDS ALSO VERY WELL TO A
CC USF/MLTF SITE. BINDING OF TFE3 TO DNA INDUCES DNA BENDING.
CC -|- SUBCELLULAR LOCATION: NUCLEAR.
CC -|- TISSUE SPECIFICITY: UBIQUITOUS IN FETAL AND ADULT TISSUES.
CC -|- DISEASE: INVOLVED IN PAPILLARY RENAL CELL CARCINOMA (PRCC) BY
CC A CHROMOSOMAL TRANSLOCATION T(X;1)(P11.2;Q21.2) WHICH INVOLVES
CC TFE3 AND PRCC.
CC -|- SIMILARITY: TO OTHER MEMBERS OF THE MYC FAMILY OF HELIX-LOOP-HELIX
CC TRANSCRIPTION FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X99721; E257625; -.
CC ENBL: X97160; E246025; -.
CC DR ENBL: X97161; E246025; JOINED.
CC DR ENBL: X97162; E246025; JOINED.
CC DR ENBL: X96717; E238023; -.
CC DR ENBL: X51330; E20907; -.
CC DR PIR: A34596; A34596.
CC PIR: S10379; S10379.


```

MTM; 314310; -.
DR DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
DR DR PFAM; PF00010; HLH; 1.
DR DR HSSP; P22415; IAN4.
DR DR TRANSFAC; T00811; -.
DR KW TRANSCRIPTION REGULATION; DNA-BINDING; ACTIVATOR; NUCLEAR PROTEIN;
KW CHROMOSOMAL TRANSLOCATION; PROTO-ONCOGENE.
FT FT DOMAIN 260 271
FT FT STRONG TRANSCRIPTION ACTIVATION DOMAIN
FT FT (POTENTIAL).
FT FT DNA_BIND 344 400
FT FT LEUCINE-ZIPPER (POTENTIAL).
FT FT DOMAIN 409 430
FT FT PRO/ARG-RICH.
FT FT SITE 575 743
FT FT BREAKPOINT FOR TRANSLLOCATION TO FORM
FT FT PRC-C-TFE3 ONCOGENE.
FT FT CONFLICT 172 172
FT FT M -> V (IN REF. 1).
FT FT CONFLICT 208 211
FT FT GPKL -> EFGK (IN REF. 3).
FT FT CONFLICT 219 219
FT FT S -> P (IN REF. 3).
FT FT CONFLICT 222 222
FT FT P -> K (IN REF. 3).
FT FT CONFLICT 229 229
FT FT L -> P (IN REF. 3).
FT FT CONFLICT 244 244
FT FT P -> A (IN E239023).
FT FT CONFLICT 443 443
FT FT P -> G (IN REF. 3).
FT FT CONFLICT 475 475
FT FT A -> R (IN REF. 3).
FT FT CONFLICT 593 593
FT FT V -> C (IN REF. 3).
FT FT CONFLICT 609 743
FT FT RYLGRDVIIRPHFCNQARRSSQMRPCTFPKGTACQRYFRRR
FT FT RLKRDMSGPAPMSQSPHPCGTHPCPGGRQDEVSTLS
FT FT PRCPQSGVWERMSGCCSLICLTGTTSLVRPGTEESSNGG
FT FT PSFSGLSAPSP -> CPGLGSEEM (IN REF. 2).
SQ SEQUENCE 743 AA; 80045 MW; 9C212765 CRC32;

Query Match 74.6%; Score 47; DB 1; Length 743;
Best Local Similarity 66.7%; Pred. No. 3.08e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 LPESGIVAD 52
QY 598 LPEDGIMAE 606
||||:

RESULT 10
ID YHJC_ECOLI STANDARD; PRT; 299 AA.
AC P37641;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN TREF-KDGK INTERGENIC REGION.
GN YHJC.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
PRT
11
SEQUENCE FROM N.A.
RC STRAIN-K12 / MGL655;
RX MEDLINE; 94315500.
RA SOFIA H.J., BURLAND V., DANIELS D.L., PLUNKETT G. III, BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL NUCLEIC ACIDS RES. 22:2576-2586(1994).
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00039; G466658; ALT_INIT
CC EMBL; AE000428; G1789938; ALT_INIT.
CC ECOGENE; EG12247; YHJC.
CC PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
CC PFAM; PF00126; HTH_1; 1.

```

```

KW  HYPOTHETICAL PROTEIN; TRANSCRIPTION REGULATION; DNA-BINDING.
FT  DNA_BIND 19 38 H-T-H MOTIF (BY SIMILARITY).
SQ  SEQUENCE 299 AA; 33329 MW; B70ECA8F CRC32;

Query Match 73.0%; Score 46; DB 1; Length 299;
Best Local Similarity 75.0%; Pred. No. 5.31e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 148 LPEDGVIA 155
|||||:|
Qy 598 LPEDGIMA 605

RESULT 11
ID ID PPV DROME STANDARD; PRT; 314 AA.
AC P11612;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE SERINE/THREONINE PROTEIN PHOSPHATASE PP-Y (EC 3.1.3.16).
GN PPV-55A.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC CEROPODILIDAE; DROSOPHILA.
[1]
RN RN SEQUENCE FROM N.A.
RP RP TISSUE-HEAD.
RC MEDLINE; 89232161.
RA DOMBRADI V., AXTON J.M., GLOVER D.M., COHEN P.T.W.;
RT "Molecular cloning and chromosomal localization of a novel Drosophila
RT protein phosphatase.";
RL FEBS LETT. 247:391-395(1989).
[2]
RN RN SEQUENCE FROM N.A.
RX RX MEDLINE; 90346193.
RA COHEN P.T.W., BREWIS N.D., HUGHES V., MANN D.J.;
RT "Protein serine/threonine phosphatases; an expanding family.";
RL FEBS LETT. 268:355-359(1990).
CC -1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O -> A PROTEIN +
CC ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
CC -1- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-Y
CC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; Y07510; G8372; -.
CC PIR; S03963; PAFFY.
CC FLYBASE; FBgn0003140; PPV-55A.
CC PROSITE; PS00125; SER_THR_PHOSPHATASE; 1.
CC PFAM; PF00149; Stphosphatase; 1.
CC HSP; P08129; IrfJM.
CC HYDROLASE; IRON; MANGANESE.
CC METAL 60 60 IRON (BY SIMILARITY).
CC FT METAL 62 62 IRON (BY SIMILARITY).
CC FT METAL 88 88 IRON AND MANGANESE (BY SIMILARITY).
CC FT METAL 120 120 MANGANESE (BY SIMILARITY).
CC FT ACT_SITE 121 121 GENERAL ACID (BY SIMILARITY).
CC FT METAL 169 169 MANGANESE (BY SIMILARITY).
CC FT METAL 244 244 MANGANESE (BY SIMILARITY).
CC SEQUENCE 314 AA; 36022 MW; 7C6C2772 CRC32;

Query Match 73.0%; Score 46; DB 1; Length 314;
Best Local Similarity 44.4%; Pred. No. 5.31e+00;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 191 IPDEGINCD 199

```



```

DR PIR: A40956; A40956.
DR PIR: A39386; A39386.
DR PDB: 1ITN: 30-APR-94.
DR PDB: 2IRT: 13-OCT-94.
DR PDB: 1IRP: 27-FEB-95.
DR PDB: 1ILR: 07-FEB-95.
DR PDB: 1ILT: 01-APR-95.
DR PDB: 1IRA: 17-JUN-98.
DR AARHUS/GHENT-2DPAGE: 7104; IEF.
DR AARHUS/GHENT-2DPAGE: 7105; IEF.
DR MIN: 147679; -.
DR PROSITE; PS00253; INTERLEUKIN_1; 1.
DR PFAM; PF00340; interleukin-1; 1.
DR GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING; 3D-STRUCTURE.
FT SIGNAL 1 25
FT CHAIN 26 177 INTERLEUKIN-1 RECEPTOR ANTAGONIST
FT DISULFID 91 141
FT CARBOHYD 109 109
FT VARSPLIC 1 21 MEICRGLRSHLITLLFLFHS -> MAL (IN
INTRACELLULAR FORM).
SQ SEQUENCE 177 AA; 20055 MW; AB5E4872 CRC32;

Query Match 71.4%; Score 45; DB 1; Length 177;
Best Local Similarity 42.9%; Pred. No. 9.09e+00;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 161 MPDEGVM 167
QY 598 LPEDGIM 604

RESULT 13
ID CFAL_MCTU STANDARD; PRT; 287 AA.
AC Q1195;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE (CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SYNTHASE 1 (EC 2.1.1.79)
DE MYCOLIC ACID SYNTHASE 1).
GN CMAA1 OR CMA1 OR RV3392C OR MTV004.50.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIAE; MYCOBACTERIUM.
OC ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RA;
RX MEDLINE; 95327699.
RA YUAN Y., LEE R.E., BESPA G.S., BELISLE J.T., BARRY C.E. III;
"Identification of a gene involved in the biosynthesis of
RT cyclopropanated mycolic acids in Mycobacterium tuberculosis.";
RL PROC. NATL. ACAD. SCI. U.S.A. 92:6630-6634(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 98295987.
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES K., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MORPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SOARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL NATURE 393:537-544(1998).
CC -!- FUNCTION: TRANSFERS A METHYLENE GROUP FROM S-ADENOSYL-L-METHIONINE
TO THE CIS DOUBLE BOND OF AN UNSATURATED FATTY ACID CHAIN
RESULTING IN THE REPLACEMENT OF THE DOUBLE BOND WITH A METHYLENE
BRIDGE. MYCOLIC ACIDS, WHICH REPRESENT THE MAJOR CONSTITUENT OF
MYCOBACTERIAL CELL WALL COMPLEX, ACT AS SUBSTRATES.

```

```

CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + PHOSPHOLIPID
CC OLEFINIC FATTY ACID -> S-ADENOSYL-L-HOMOCYSTEINE + PHOSPHOLIPID
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U27357; G886102; -.
DR EMBL; AL009198; E1202310; -.
KW TRANSFERASE; METHYLTRANSFERASE; LIPID SYNTHESIS.
FT DOMAIN 68 76 S-ADENOSYL-L-METHIONINE-BINDING MOTIF.
FT ACT_SITE 269 269 ESSENTIAL FOR CATALYSIS (PROBABLE).
SQ SEQUENCE 287 AA; 32461 MW; F7AB4164 CRC32;

Query Match 71.4%; Score 45; DB 1; Length 287;
Best Local Similarity 71.4%; Pred. No. 9.09e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 158 LPADGVM 164
QY 598 LPEDGIM 604

RESULT 14
ID SPEE_METJA STANDARD; PRT; 293 AA.
AC Q57761;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PUTATIVE SPERMIDINE SYNTHASE (EC 2.5.1.16) (PUTRESCINE
DE AMINOPROPYLTRANSFERASE) (SPDSY).
GN MJ0313.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
OC METHANOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TORB J.F., ADAMS M.D., REICH C.I.,
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
RA SCOTT J.L., GEORHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL SCIENCE 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: S-ADENOSYLMETHIONINAMINE + PUTRESCINE -
CC 5'-METHYLTHIOADENOSINE + SPERMIDINE (BY SIMILARITY).
CC -!- PATHWAY: FIFTH (LAST) STEP IN THE BIOSYNTHESIS OF SPERMIDINE FROM
CC ARGinine AND METHIONINE (BY SIMILARITY).
CC -!- SIMILARITY: IN THE SUBSTRATE-BINDING REGION, TO SPERMINE SYNTHASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67486; G1591033; -.
DR TIGR; MJ0313; -.
KW HYPOTHETICAL PROTEIN; SPERMIDINE BIOSYNTHESIS; TRANSFERASE.

```

FT DOMAIN 85 127 BINDING TO DECARBOXYLATED SAM
FT (POTENTIAL).
SQ SEQUENCE 293 AA; 33899 MW; 538CB8D8 CRC32;

Query Match 71.4%; Score 45; DB 1; Length 293;
Best Local Similarity 55.6%; Pred. No. 9.09e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 187 LNDDGIMVQ 195

|:||||:

QY 598 LPEDGIMAE 606

RESULT 15

ID RIBG_BACAM STANDARD; PRT; 371 AA.
AC P70814;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE RIBOFLAVIN-SPECIFIC DEAMINASE (EC 3.5.4.-).
GN RIBG.
OS BACILLUS AMYLOLIQUEFACIENS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A 50;
RA GUSAROV I. I.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -!- FUNCTION: INVOLVED IN RIBOFLAVIN BIOSYNTHESIS. CONVERTS
CC 2,5-DIAMINO-6-(RIBOSYLAMINO)-4(3H)-PYRIMIDINONE 5'-PHOSPHATE
CC INTO 5-AMINO-6-(RIBOSYLAMINO)-2,4(1H,3H)-PYRIMIDINEDIONE
CC 5'-PHOSPHATE (BY SIMILARITY).
CC -!- COFACTOR: ZINC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY. STRONG, TO OTHER RIBOFLAVIN-SPECIFIC DEAMINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: X95955; E270077;
DR PROSITE: PS00903; CYT_DCMP_DEAMINASES; 1.
DR PFAM: PF00383; dCMP_cyt_deam; 1.
KW RIBOFLAVIN BIOSYNTHESIS; HYDROLASE; ZINC.
FT METAL 49 49 ZINC (BY SIMILARITY).
FT METAL 74 74 ZINC (BY SIMILARITY).
FT METAL 83 83 ZINC (BY SIMILARITY).
SQ SEQUENCE 371 AA; 40166 MW; 36960A38 CRC32;

Query Match 71.4%; Score 45; DB 1; Length 371;
Best Local Similarity 62.5%; Pred. No. 9.09e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 279 LAENGIMS 286

|:||||:

QY 598 LPEDGIMA 605

Search completed: Thu Jul 8 19:08:21 1999

Job time : 8 secs.

WQISREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:08:40 1999; MasPar time 7.61 Seconds
64.574 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (598-606) from US09041236.pep (30 of 45)
Perfect Score: 63
Sequence: 1 LPEDGIMAE 9

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrmb19
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 22.091; Variance 24.334; scale 0.908

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	63	100.0	666	4	075326 SEMAPHORIN L.	8.22e-04
2	52	82.5	258	2	069165 ACCG'.	5.11e-01
3	51	81.0	494	4	060849 HYPOTHETICAL 54.2 KD P	8.86e-01
4	48	76.2	272	5	018776 SIMILAR TO POTASSIUM C	4.42e+00
5	46	73.0	258	2	005664 RPON AND ORF GENES.	1.24e+01
6	46	73.0	264	10	049481 HYPOTHETICAL 30.8 KD P	1.24e+01
7	46	73.0	604	2	005617 BIOTIN CARBOXYL CARRIE	1.24e+01
8	46	73.0	704	10	096268 HEAT SHOCK PROTEIN 90A	1.24e+01
9	45	71.4	180	4	014628 INTRACELLULAR IL-1 REC	2.06e+01
10	45	71.4	200	1	057820 200AA LONG HYPOTHETICA	2.06e+01
11	45	71.4	259	2	069441 29 KDA IMMUNOGENIC PRO	2.06e+01
12	45	71.4	342	14	072486 ENVELOPE GLYCOPROTEIN	2.06e+01
13	45	71.4	351	14	087978 ENVELOPE GLYCOPROTEIN	2.06e+01
14	45	71.4	416	6	027984 ALPHA1-ANTICHYMYOTRPSI	2.06e+01
15	45	71.4	450	14	072487 ENVELOPE GLYCOPROTEIN	2.06e+01
16	45	71.4	451	14	072485 ENVELOPE GLYCOPROTEIN	2.06e+01
17	45	71.4	1503	2	052973 CYCLOINULO-OLIGOSACCHA	2.06e+01
18	45	71.4	1510	5	061802 H1e01.3 PROTEIN.	2.06e+01
19	44	69.8	208	14	074044 ENVELOPE GLYCOPROTEIN,	3.39e+01
20	44	69.8	242	1	026333 CONSERVED PROTEIN.	3.39e+01

21	44	69.8	310	14	086551 POL-POLYMERASE, E2-PEP	3.39e+01
22	44	69.8	443	2	P76507 FROM BASES 2458322 TO	3.39e+01
23	44	69.8	687	3	006147 CHROMOSOME XII COSMID	3.39e+01
24	44	69.8	1026	11	062845 NEURAL CELL ADHESION P	3.39e+01
25	44	69.8	1099	11	P97527 KB-2.	3.39e+01
26	44	69.8	1522	4	015069 KIAA0363 (FRAGMENT).	3.39e+01
27	44	69.8	1932	3	006409 CHROMOSOME XII COSMID	3.39e+01
28	43	68.3	185	2	086434 PUTATIVE ACETYL TRANSF	5.53e+01
29	43	68.3	203	6	027983 ALPHA1-ANTICHYMYOTRPSI	5.53e+01
30	43	68.3	252	6	028921 ALPHA 1-ANTICHYMYOTRPS	5.53e+01
31	43	68.3	334	10	004695 DNA-BINDING PDI-LIKE P	5.53e+01
32	43	68.3	339	2	035045 ZINC-CONTAINING ALCOHO	5.53e+01
33	43	68.3	347	10	004696 DNA-BINDING PROTEIN PD	5.53e+01
34	43	68.3	383	2	031028 ASPARTATE AMINOTRANSE	5.53e+01
35	43	68.3	383	2	038222 ASPARTATE AMINOTRANSE	5.53e+01
36	43	68.3	385	2	056232 ASPARTATE AMINOTRANSE	5.53e+01
37	43	68.3	423	11	P97569 KALLISTATIN.	5.53e+01
38	43	68.3	487	14	071628 ENVELOPE GLYCOPROTEIN	5.53e+01
39	43	68.3	517	14	074848 ENVELOPE GLYCOPROTEIN	5.53e+01
40	43	68.3	582	2	067141 ELONGATION FACTOR SELB	5.53e+01
41	43	68.3	696	5	Q17034 IMMUNE FACTOR.	5.53e+01
42	43	68.3	890	10	023009 HYPOTHETICAL 98.8 KD P	5.53e+01
43	43	68.3	1043	2	Q51396 RND FAMILY EXPORTER ME	5.53e+01
44	43	68.3	1890	10	049528 GENE 11-1 PROTEIN - LI	5.53e+01
45	43	68.3	8563	2	Q54297 POLYKETIDE SYNTHASE.	5.53e+01

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT; 666 AA.
ID 075326		
AC 075326:		
DT 01-NOV-1998 (TREMBREL. 08, CREATED)		
DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)		
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)		
DE SEMAPHORIN L.		
GN SEMAL		
OS HOMO SAPIENS (HUMAN).		
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;		
OC CATARRHINI; HOMINIDAE; HOMO.		
RN [1]		
RP SEQUENCE FROM N.A.		
RX MEDLINE; 98399619.		
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;		
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA		
RT viruses."		
RL GENOMICS 51:340-350(1998).		
DR EMBL; AF030698; G3523115; -.		
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;		

Query Match 100.0%; Score 63; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 8.22e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	630 LPEDGIMAE 638
Qy	598 LPEDGIMAE 606

RESULT 2
ID 069165
AC 069165:

DT 01-AUG-1998 (TREMBREL. 07, CREATED)	
DT 01-AUG-1998 (TREMBREL. 07, LAST SEQUENCE UPDATE)	
DT 01-AUG-1998 (TREMBREL. 07, LAST ANNOTATION UPDATE)	
DE ACCG'.	
GN ACCG'.	
OS AGROBACTERIUM TUMEFACIENS.	
OG PLASMID PTIC58.	
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;	
OC RHIZOBIACEAE; AGROBACTERIUM.	
RN [1]	
RP SEQUENCE FROM N.A.	

RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RX	MEDLINE; 94150718.
RA	WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA	BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA	CRAWTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA	JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA	LIGHTNING J., LLOYD C., McMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA	PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA	SALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA	THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT	elegans";
RL	NATURE 368:32-38(1994).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RA	MARTIN J.;
RL	SUBMITTED (DEC-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RA	WATERSTON R.;
RL	SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR	EMBL; U41276; G1086886; "
KW	IONIC CHANNEL
SQ	SEQUENCE 272 AA; 30724 MW; 478DSDD92 CRC32;
Dd	Query Match 76.2%; Score 48; DB 5; Length 272;
	Best Local Similarity 66.7%; Pred. No. 4.42e+00;
	Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps
Qy	155 LSEGGILAE 163 :: :: 598 LPEDGIMAE 606
RESULT 5	
ID	O05664 PRELIMINARY; PRT: 258 AA.
AC	O05664;
DT	01-JUL-1997 (TREMBLREL. 04, CREATED)
DT	01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DE	01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE	RPN AND ORF GENES.
OS	LISTERIA MONOCYTOGENES.
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
CC	LISTERIA.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=lo28 SEROVAR 1/2C;
RA	ROBICHON D.G.E.;
RL	SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR	EMBL; X93169; E213288; "
SQ	SEQUENCE 258 AA; 29041 MW; 9FC46B50 CRC32;
Dd	Query Match 73.0%; Score 46; DB 2; Length 258;
	Best Local Similarity 66.7%; Pred. No. 1.24e+01;
	Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps
Qy	87 LAADGIME 95 :: 598 LPEDGIMAE 606
RESULT 6	
ID	O49481 PRELIMINARY; PRT: 264 AA.
AC	O49481;
DT	01-JUN-1998 (TREMBLREL. 06, CREATED)
DT	01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE	01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE	HYPOTHETICAL 30.8 KD PROTEIN.

GN F24J7.200.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., VITALE D., LIGUORI R., ARGIRIOU A., DE SIMONE V.,
 RA BANCROFT I., MEWES H.W., MAYER K., SCHUELLER C.;
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT;
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AL021768; E1250355; -
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 264 AA; 30818 MW; 3F53ABC8 CRC32;

 Query Match 73.0%; Score 46; DB 10; Length 264;
 Best Local Similarity 66.7%; Pred. No. 1.24e+01;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 Db 155 LSDGIMFE 163
 I:|||||
 QY 598 LPEDGIMAE 606

 RESULT 7
 ID Q05617 PRELIMINARY; PRT; 604 AA.
 AC Q05617;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE BIOTIN CARBOXYL CARRIER PROTEIN OF METHYLMALONYL-COA
 DE CARBOXYL-TRANSFERASE (EC 2.1.3.1) (TRANSCARBOXYLASE, 12S SUBUNIT).
 OS PROPIONIBACTERIUM FREUDENEICHI SHERMANII.
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
 OC ACTINOMYCETALES; PROPIONIBACTERINEAE; PROPIONIBACTERIACEAE;
 OC PROPIONIBACTERIUM.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-ST33;
 RX MEDLINE: 93374821.
 RA THORNTON C.G., KUMAR G.K., HAASE C.F., WOO S.B., PARK V.M.,
 RA MAGNER W.J., WOOD H.G., SAMOULS D.;
 RT "Primary structure of the monomer of the 12S subunit of
 RT transcarboxylase as deduced from DNA and characterization of the
 RT product expressed in *Escherichia coli*.";
 RL J. BACTERIOL. 175:5301-5308(1993).
 CC -1- FUNCTION: THE 12S SUBUNIT CATALYZES THE TRANSFER OF THE CARBOXYL
 CC GROUP OF METHYLMALONYL COENZYME A TO THE BIOTIN OF THE 1,3S
 CC SUBUNIT.
 CC -1- CATALYTIC ACTIVITY:
 CC (S)-2-METHYL-3-OXOPROPANOL-COA + PYRUVATE -> PROPANOL-COA +
 CC OXALACETATE.
 CC -1- SUBUNIT: HOMOHETEROMER.
 CC -1- SUBUNIT: TRANSCARBOXYLASE IS COMPOSED OF THREE SUBUNITS: 1.3S,
 CC 5S, AND 12S.
 CC THE CORE OF THE ENZYME IS COMPOSED OF SIX 12S SUBUNITS.
 CC ON EACH SIDE OF THE CORE THERE ARE THREE PAIRS OF 5S SUBUNITS.
 CC EACH 5S DIMER IS ATTACHED TO THE CORE BY TWO 1.3S SUBUNITS.
 CC THUS THE TOTAL NUMBER OF CHAINS IS 30 (6 + 12 + 12).
 CC -1- COFACTOR: BIOTIN.
 CC -1- SIMILARITY: REGIONALLY, TO PROPIONYL-COA CARBOXYLASE BETA CHAIN
 CC (P05166) FROM RAT (P07633), PCCASE FROM HUMAN
 CC (P05166) AND TO ACETYL-COENZYME A CARBOXYLASE BETA SUBUNIT FROM
 CC *E. COLI* (P08193).
 DR EMBL; L04196; G150933; -
 DR PFAM; PF01039; Carboxyl_trans; 1.
 KW TRANSFERASE.
 FT INIT_MET 0
 SQ SEQUENCE 604 AA; 65568 MW; CBE86FEA CRC32;

Query Match 73.0%; Score 46; DB 2; Length 604;
 Best Local Similarity 55.6%; Pred. No. 1.24e+01;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

 Db 518 LSEEEIMAD 526
 I:|||||
 QY 598 LPEDGIMAE 606

 RESULT 8
 ID Q96268 PRELIMINARY; PRT; 704 AA.
 AC Q96268;
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE HEAT SHOCK PROTEIN 90A.
 GN HSP90A.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98088020.
 RA MILIONI D., HATZOPOULOS P.;
 RT "Genomic organization of hsp90 gene family in *Arabidopsis*.";
 RL PLANT MOL. BIOL. 35:955-961(1997).
 DR EMBL; Y07613; E263435; -
 DR PROSITE; PS00298; HSP90; 1.
 DR PFAM; PF00183; HSP90; 1.
 DR MENDEL; 7102; ARATH; 1518; 3.
 KW HEAT SHOCK.
 SQ SEQUENCE 704 AA; 81063 MW; 6BA64FD5 CRC32;

 Query Match 73.0%; Score 46; DB 10; Length 704;
 Best Local Similarity 50.0%; Pred. No. 1.24e+01;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

 Db 615 PDNGIMED 622
 I:|||||
 QY 599 PEDGIMAE 606

 RESULT 9
 ID Q14628 PRELIMINARY; PRT; 180 AA.
 AC Q14628;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE INTRACELLULAR IL-1 RECEPTOR ANTAGONIST TYPE II.
 GN IL-1RN.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95355865.
 RA MUZIO M., PLENTARUTTI N., SIRONI M., POLI G., DE GIOIA L.,
 RA INTRONA M., MANTOVANI A., COLOTTA F.;
 RT "Cloning and characterization of a new isoform of the interleukin 1
 RT receptor antagonist.";
 RL J. EXP. MED. 182:623-628(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97146044.
 RA JENKINS J.K., DRONG R.F., SHUCK M.E., BIENKOWSKI M.J., SLIGHTOM J.L.,
 RA AREND W.P., SMITH M.F. JR.;
 RT "Intracellular IL-1 receptor antagonist promoter: cell type-specific
 RT and inducible regulatory regions.";
 RL J. IMMUNOL. 158:748-755(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX SLIGHTOM J.L.;

RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: X84348; G1008971; -
DR EMBL: U65590; G2707376; -
DR PROSITE: PS00253; INTERLEUKIN_1; 1.
DR PFAM: PF00340; interleukin-1; 1.
SQ SEQUENCE 180 AA; 19897 MW; 111E8393 CRC32;

Query Match 71.4%; Score 45; DB 4; Length 180;
Best Local Similarity 42.9%; Pred. No. 2.06e+01;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 164 MPDEGVM 170
QY 598 LPEDGIM 604

RESULT 10
ID O57820 PRELIMINARY; PRT: 200 AA.
AC O57820;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE 200AA LONG HYPOTHETICAL PROTEIN.
GN PH0043.
OS PYROCOCCLUS HORIKOSHII.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOVANA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete Sequence and Gene Organization of the Genome of a
RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 5:55-76(1998).
DR EMBL: AP000001; D1030054; -
SQ SEQUENCE 200 AA; 22277 MW; C00C616F CRC32;

Query Match 71.4%; Score 45; DB 1; Length 200;
Best Local Similarity 66.7%; Pred. No. 2.06e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 30 LPDGVIAE 38
QY 598 LPEDGIMAE 606

RESULT 11
ID O69441 PRELIMINARY; PRT: 259 AA.
AC O69441;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE 29 KDA IMMUNOGENIC PROTEIN.
OS LEGIONELLA PNEUMOPHILA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; LEGIONELLACEAE;
OC LEGIONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CORBY;
RA STEUDEL C., HELBIG J., LUECK C.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AJ005668; E1288207; -
SQ SEQUENCE 259 AA; 28614 MW; 004D3AB0 CRC32;

Query Match 71.4%; Score 45; DB 2; Length 259;
Best Local Similarity 75.0%; Pred. No. 2.06e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 119 LPENGIIA 126
QY 598 LPEDGIMA 605

RESULT 12
ID Q72486 PRELIMINARY; PRT: 342 AA.
AC Q72486;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=94-47622, 1;
RX MEDLINE; 96020103.
RA ARNOLD C., BARLOW K.L., KAYE S., LOVEDAY C., BALFE P., CLEWLEY J.P.;
RT "HIV type 1 sequence subtype G transmission from mother to infant:
RT failure of variant sequence species to amplify in the Roche Amplior
RT Test.";
RL AIDS RES. HUM. RETROVIRUSES 11:999-1001(1995).
DR EMBL: U26305; G852514; -
DR PFAM: PF00516; GP120; 1.
KW ENVELOPE PROTEIN.
FT NON_TER 1 1
FT NON_TER 342 342
SQ SEQUENCE 342 AA; 37963 MW; 9632142C CRC32;

Query Match 71.4%; Score 45; DB 14; Length 342;
Best Local Similarity 71.4%; Pred. No. 2.06e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 142 LAEGGIM 148
QY 598 LPEDGIM 604

RESULT 13
ID P87978 PRELIMINARY; PRT: 351 AA.
AC P87978;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN, VI-V5 REGION (FRAGMENT).
GN ENV.
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA XIN K.Q., SHAPSHAK P.S.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA XIN K.Q., SHAPSHAK P.S., OKUDA K., BRADLEY W.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U25258; G1754957; -
DR PFAM: PF00516; GP120; 1.
KW ENVELOPE PROTEIN.
FT NON_TER 1 1
FT NON_TER 351 351
SQ SEQUENCE 351 AA; 38605 MW; ED60457F CRC32;

Query Match 71.4%; Score 45; DB 14; Length 351;
Best Local Similarity 71.4%; Pred. No. 2.06e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 146 LSEEGIM 152
QY 598 LPEDGIM 604


```
RESULT 14
ID Q27984 PRELIMINARY; PRT; 416 AA.
AC Q27984;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ALPHAI-ANTICHYOTRYPSIN ISOFORM PHK12 (FRAGMENT).
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95023955.
RA HWANG S.R., KOHN A.B., HOOK V.Y.;
RT "Molecular cloning reveals isoforms of bovine alpha
RL 1-antichymotrypsin.";
RL PROC. NATL. ACAD. SCI. U.S.A. 91:9579-9583(1994).
DR EMBL: U13609; G53509; -.
DR PFAM: PF00079; serpin; 1.
FT NON_TER 1
SQ SEQUENCE 416 AA; 46953 MW; 46FBAE02 CRC32;

Query Match 71.4%; Score 45; DB 6; Length 416;
Best Local Similarity 44.4%; Pred. No. 2.06e+01;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 273 LPDEGKMOD 281
|::|:
QY 598 LPEDGINAE 606

RESULT 15
ID Q72487 PRELIMINARY; PRT; 450 AA.
AC Q72487;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=94-47622; 2;
RX MEDLINE: 96020103.
RA ARNOLD C., BARLOW K.L., KAYE S., LOVEDAY C., BALFE P., CLEWLEY J.P.;
RT "HIV type 1 sequence subtype G transmission from mother to infant:
RT failure of variant sequence species to amplify in the Roche Amplicor
RT Test.";
RL AIDS RES. HUM. RETROVIRUSES 11:999-1001(1995).
DR EMBL: U26306; G852516; -.
DR PFAM: PF00516; GP120; 1.
KW ENVELOPE PROTEIN.
FT NON_TER 1
FT NON_TER 450
SQ SEQUENCE 450 AA; 50528 MW; 2ECFDC76 CRC32;

Query Match 71.4%; Score 45; DB 14; Length 450;
Best Local Similarity 71.4%; Pred. No. 2.06e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 230 LAEEGIM 236
|::|:
QY 598 LPEDGIM 604
```

Search completed: Thu Jul 8 19:08:53 1999
Job time : 13 secs.

THIS PAGE BLANK (USPTO)

WQSEH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:13:03 1999; MasPar time 30.37 Seconds
Tabular output not generated. 419.473 Million cell updates/sec

Title: >US-09-041-236-2

Description: (8-606) from US09041236.pap (31 of 45)

Perfect Score: 4461
Sequence: 1 AAASAOGLRSGPRIFAVWK.....SYFRAQHOLLPEGIMAE 599

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.527; Variance 154.407; scale 0.237

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	661	14.8	771	13	Human semaphorin III	6.82e-51
2	620	13.9	477	13	Human collapsin.	6.84e-47
3	608	13.6	775	33	Human semaphorin.	1.01e-45
4	606	13.6	861	22	Mouse CD100 antigen.	1.58e-45
5	599	13.4	861	32	Human semaphorin.	7.57e-45
6	544	12.2	862	22	Human CD100 antigen.	1.63e-39
7	518	11.6	776	32	Rat semaphorin W.	5.28e-37
8	463	10.4	441	13	Vaccinia virus semaph	1.01e-31
9	459	10.3	730	13	Grasshopper semaphori	2.44e-31
10	428	9.6	587	32	Human semaphorin W.	2.20e-28
11	390	8.7	974	33	Human secreted protei	1.18e-25
12	378	8.5	712	13	Tribolium semaphorin	1.78e-23
13	367	8.2	650	13	Drosophila semaphorin	1.27e-22
14	362	8.1	724	13	Drosophila semaphorin	3.74e-22
15	352	7.9	930	32	Human semaphorin Y.	3.23e-21
16	325	7.3	888	25	Human semaphorin Z.	1.05e-18

17	316	7.1	887	25	W19856	Rat semaphorin Z.	7.15e-18
18	296	6.6	929	32	W57259	Rat semaphorin Y.	4.96e-16
19	157	3.5	122	13	R71385	Variola major virus s	8.30e-04
20	101	2.3	2329	25	W25038	Partial BRCA2 cancer	1.86e-01
21	99	2.2	591	4	R23006	Protein transcribed f	2.58e-01
22	98	2.2	832	33	W61092	Taq DNA polymerase I	3.04e-01
23	96	2.2	832	33	W61090	Taq DNA polymerase I	4.20e-01
24	92	2.1	832	33	W61091	Taq DNA polymerase I	7.96e-01
25	92	2.1	854	30	W56309	Clas II S-receptor ki	7.96e-01
26	92	2.1	855	6	R29815	S receptor kinase pro	7.96e-01
27	93	2.1	4572	30	W52845	A. mediterranei rifam	6.79e-01
28	91	2.0	478	1	R04881	Recombinant elastase	9.33e-01
29	89	2.0	534	25	W25031	Partial BRCA2 cancer	1.28e-02
30	89	2.0	554	13	R66209	Novel thermostable DN	1.28e-02
31	91	2.0	598	9	R48631	Sequence of nuclear r	9.33e-01
32	90	2.0	638	5	R03924	E. coli HSP (dnaK).	1.09e-02
33	89	2.0	680	4	R23143	Mutant thermostable D	1.28e-02
34	89	2.0	832	13	R76693	DNA-polymerase (F73L,	1.28e-02
35	89	2.0	832	1	P90556	Purified native therm	1.28e-02
36	89	2.0	832	33	W61087	Taq DNA polymerase I	1.28e-02
37	89	2.0	832	33	W61088	Taq DNA polymerase I	1.28e-02
38	89	2.0	832	13	R76690	Taq DNA-polymerase RE	1.28e-02
39	89	2.0	832	33	W61089	Taq DNA-polymerase I	1.28e-02
40	89	2.0	833	36	W59942	Amino acid sequence o	1.28e-02
41	89	2.0	833	27	W24211	Cleavase DN nuclease.	1.28e-02
42	89	2.0	833	27	W24212	Cleavase DA nuclease.	1.28e-02
43	90	2.0	1242	30	W52287	Rattus norvegicus cdo	1.09e-02
44	89	2.0	3418	26	W19211	Human breast cancer s	1.28e-02
45	89	2.0	3418	26	W23287	Human breast and ovar	1.28e-02

ALIGNMENTS

RESULT 1
ID R71380 standard; Protein; 771 AA.

AC R71380;
DT 21-NOV-1995 (first entry)
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW Variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN WO9507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB: Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 60-63; 101pp; English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II,
CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor binding
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA;

Query Match 14.8%; Score 661; DB 13; Length 771;

Best Local Similarity 30.3%; Pred. No. 6.82e-51;

Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;

Db 166 rgkspydpklltasllidgelysgta-adfmgdrfaftlghhhpirtqehdsrwndp 224

Db 343 vyhmssireafngpyahkegpyhwslyegkvyprpbgscaskvnggkygttktdypddai 402
 Qy 304 VYSLGDIDKVRTS-SLK-G--YHSSL-----PNRPGKCLPD-QOPI--PTETP-QVAD 350
 Db 403 rfarnhplmyqpkvhhkpllvktdgkynlqlavdrveaedgg-ydvlfigttdgvl 461
 Qy 351 RIPEVAQRV-EPWGPL-KTP-LFHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTH 405
 Db 462 kvityngtetweevleelqfkdapilsmesskrqqlgisasavaqvrfhcdm 521
 Qy 406 KVEPEGEQSHSAFN-IME-IQPFRAAAIQTMSLDAERRKLYVSSQWEVSQVPLDCEV 463
 Db 522 ygsacacclardpycaudgiscsryypt 550
 Qy 464 YGGGCHGCLMSRDPYCGWDQGRICSIYSS 492

RESULT 4

ID W17658 standard; Protein; 861 AA.
 AC W17658;
 DT 24-JUL-1997 (first entry)
 DE Mouse CD100 antigen.
 KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
 KW vaccine.
 OS Mus sp.
 FH key
 FT peptide
 FT protein
 FT domain
 FT domain
 FT domain
 FT domain
 FT domain
 FT domain
 FT domain
 FT modified_site
 FT
 FT W09717368-A1.
 PN 15-MAY-1997.
 PD 12-NOV-1996; U18645.
 PR 09-NOV-1995; US-556422.
 PA (DAND) DANA FARBEN CANCER INST.
 PI Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
 DR WPI; 97-280982/25.
 DR N-PSDB; T60666.
 PT Nucleic acid molecule encoding CD100 antigen - which stimulates
 PT leukocyte response, e.g. B cell aggregation, differentiation,
 PT survival and T cell proliferation
 PS Example 8; Page 86-89; 135pp; English.
 CC Mouse CD100 antigen (W17657) is a novel leukocyte semaphorin-like
 CC protein that stimulates a leukocyte response, including B cell
 CC aggregation, B cell differentiation, B cell survival and/or T cell
 CC proliferation. Its amino acid sequence was deduced from a cDNA
 CC clone (T60666) isolated from murine T cells. Human CD100 antigen
 CC (W17657) has also been identified. CD100 polypeptides and fusion
 CC proteins, nucleic acids, and host cells expressing CD100 can be
 CC utilised in diagnostic and therapeutic methods involving modulation
 CC of B and T cell responses, neuron axonal growth and immune cell-
 CC nerve cell interaction.
 SQ Sequence 861 AA;

Query Match 13.6%; Score 606; DB 22; Length 861;
 Best Local Similarity 32.4%; Pred. No. 1.58e-45;
 Matches 143; Conservative 106; Mismatches 145; Indels 47; Gaps 31;

Db 107 eclnyrvlqlsstsllyvcgtnafqptcdhnltsfkflgksedgkgrcfdpahsyts 166

Qy 93 DCENITILLER-RSEGLACGINARHPSG--WNLVNGTVVPLGEM-RGVAPSPDENSIV 148
 Db 167 vmvggelysgts-yfnlgsepiisnshsplrteyaipwlnepsvfadvlqkspdgpe 225
 Qy 149 LFEDEVYSTIRKQEVNGKIPREFRIRGESELYTSDTV--MQNPOFIKATIVHQ--D--Q 202
 Db 226 geddkvyyfftevsveyefvklmiprvavrcvkgdggllrtlq-kkwtseflkarlicskp 284
 Qy 203 AYDDKIYFFREDNDPKNPEAPLNVSRVAQLCRGQGGESSLSVSKWNTFLKAMLVCSDA 262
 Db 285 dsglvfnlqdvfvlrap-gl-kepvfyavftpqlnnvglsvacaytlatveavfsrgky 342
 Qy 263 ATKNFNRLQDVFLPDPSGGWRDTRVGVFSNPWNY---SACVYSLGDIKVP-R--- 315
 Db 343 mgsatveqshktwryngvpptprpgacidsearaanytsslnlpdktlqfkdhlpmdd 402
 Qy 316 --TSSL----K-G-YHSSLPNRPKGKCL-PDQQPIP-TETQVADRHPE-V-AQR-VE- 360
 Db 403 svtpidnprkllkdvnytiqivvdrtdqldgtfyvmfistdrgalhhkavlltkveh--v 460
 Qy 361 PMGPLKT-P-LFHSKYHYQKVAVHRMQASHGETFHVLYLTDRGTHKVVPEGEHSPA 418
 Db 461 --ieetqlfrdsepyvltlllsskkgrkfvyagsnvgvqaplafcekbgs-cedcvlard 517
 Qy 419 FNIMEIQPFRAAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDCEVYGGCHGCLMSRD 476
 Db 518 pycawspaikacvtlqhgeas 538
 Qy 477 PYCGWDQG-R-CISIIYSSERS 495

RESULT 5
 ID W58540 standard; Protein; 861 AA.
 AC W58540;
 DT 02-SEP-1998 (first entry)
 DE Human semaphorin.
 KW Human; semaphorin; diagnosis; nervous disease; immune disease;
 KW nerve extension inhibiting activity.
 OS Homo sapiens.
 PN J10155490-A.
 PD 16-JUN-1998.
 PR 27-NOV-1996; 332900.
 PR 27-NOV-1996; JP-332900.
 PA (SUMU) SUMITOMO SEIYAKU KK.
 DR WPI; 98-391044/34.
 DR N-PSDB; V31121.
 PT New human semaphorin gene - useful in the diagnosis of nervous system
 PT and immune disorders
 PS Claim 1; Page 10-12; 15pp; Japanese.
 CC The present sequence represents human semaphorin (translated from the
 CC Japanese specification as semaforin). Semaphorin has nerve extension
 CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
 CC treatment and researches on nervous diseases and immune diseases.
 SQ Sequence 861 AA;

Query Match 13.4%; Score 599; DB 32; Length 861;
 Best Local Similarity 32.4%; Pred. No. 7.57e-45;
 Matches 143; Conservative 105; Mismatches 146; Indels 47; Gaps 31;

Db 107 eclnyrvlqlsstsllyvcgtnafqptcdhnltsfkflgksedgkgrcfdpahsyts 166
 Qy 93 DCENITILLER-RSEGLACGINARHPSG--WNLVNGTVVPLGEM-RGVAPSPDENSIV 148
 Db 167 vmvggelysgts-yfnlgsepiisnshsplrteyaipwlnepsvfadvlqkspdgpe 225
 Qy 149 LFEDEVYSTIRKQEVNGKIPREFRIRGESELYTSDTV--MQNPOFIKATIVHQ--D--Q 202

Query Match 13.6%; Score 606; DB 22; Length 861;
 Best Local Similarity 32.4%; Pred. No. 1.58e-45;
 Matches 143; Conservative 106; Mismatches 145; Indels 47; Gaps 31;

Db 226 geddkvyyfftevsveyefvklmiprvavrcvkgdggllrtlq-kkwtseflkarlicskp 284
 Qy 203 AYDDKIYFFREDNDPKNPEAPLNVSRVAQLCRGQGGESSLSVSKWNTFLKAMLVCSDA 262
 Db 285 dsglvfnlqdvfvlrap-gl-kepvfyavftpqlnnvglsvacaytlatveavfsrgky 342

Qy 263 ATKNNFNRLQDVFLLPDPSCGWRDTRVGVFSPNPNY----SVCVYSLGDDDKVF-R--- 315
Db 343 mqsatvegshkvwryngpvtprgagacidsearaanytsslnlpdktlqfvgkdhplmdd 402
Qy 316 --TSSL-----K-G-YHSSLNPNRPGKCL-PDQOPIP-TETFOVADRHPE-V-AQR-VE- 360
Db 403 svtpidnrpklikdvnyvtqivvdrtdqldgtfydvmsfistdrgalhkaviltkev--v 460
Qy 361 PMGPKLT-P-LFHSKYHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVVEFGQEHSA 418
Db 461 --ieetqlfrfepvltlllskgrkfvyagsngvvqaplafeckhgs--cedcvlard 517
Qy 419 FNIMEIOPFRRAAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDLCEVYGGCHGLMSRD 476
Db 518 pycawsapaicavtlhqeas 538
Qy 477 PYCGWDQG-R-CISYSSERS 495

RESULT

ID W17657 standard; Protein; 862 AA.
AC W17657;
DT 24-JUL-1997 (first entry)
DE Human CD100 antigen.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..41
FT protein /label= Sig_peptide
FT protein 42..862
FT domain /label= Mat_protein
FT domain 42..553
FT domain /label= Semaphorin_domain
FT domain 554..630
FT domain /label= Ig-like_domain
FT domain 631..733
FT domain /label= Stalk_domain
FT domain 735..752
FT domain /label= Transmembrane_domain
FT domain 753..862
FT modified_site /label= Cytoplasmic_domain
FT modified_site 808..815
FT modified_site /label= Phosphorylation
FT modified_site /note= "putative tyrosine phosphorylation site"
PN W09717368-A1.
PD 15-MAY-1997.
PF 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND) DANA FARBER CANCER INST.
PI Boussioltis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB: T60655.
DR Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Claim 7: Page 70-72: 135pp: English.
CC Human CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60665) isolated by COS cell expression cloning. CD100
CC polypeptides and fusion proteins, nucleic acids, and host cells
CC expressing CD100 can be utilised in diagnostic and therapeutic
CC methods involving modulation of B and T cell responses, neuron
CC axonal growth and immune cell-nerve cell interaction.
SQ Sequence 862 AA;

Query Match 12.2%; Score 544; DB 22; Length 862;

Best Local Similarity 29.3%; Pred. No. 1.63e-39;
Matches 157; Conservative 134; Mismatches 189; Indels 55; Gaps 39;

Db 107 eclynrivrlplsaatslyvcgtnafpacdhlalitsfkflgknedkgropfdpahsyts 166
Qy 93 DCENYITLLERRSE-GLACGTNARHPSCNVLNVTVPPLG--EM-RGYAFSPDEN-SL 147
Db 167 vmdvg-elysgts--ynflgseplsrnshspltarvaylpwnopsfvfadvirkspdsp 224
Qy 148 VLFEDEVGVYTIKQEVYKIPRFRIRGESELYSTV--MQNPQFIKATIVHQ--D-- 201
Db 225 dgeddryvfftevsvevfvrlipriarvcvkgdggllrtiq-kkwtstflkarlicsr 283
Qy 202 QAYDDKIIYFFREDNPKNPEAPLNWSRVAQLCKDGGESSLSVSKWTFLLKAMLVCS 261
Db 284 pdegslvfnvdrfvrlsp-gl-kvpvfyalfpqlnnvglsavcaylnlstaecvfhgk 341
Qy 262 AATNKNFNRLQDVFLLPDPSCGWRDTRVGVFSPNPNY---SVCVYSLGDDDKVF--R 315
Db 342 ymsttveqshkvwryngpvtprgagacidsearaanytsslnlpdktlqfvgkdhplm 401
Qy 316 ----TS---S-LK-G-YHSSLNPNRPGKCL-PDQOPIP-TETFOVADRHPE-V-AQR-VE 360
Db 402 dsvtpidnrprlikdvnyvtqivvdrtdqldgtvydvmsfistdrgalhkai--s-lehav 458
Qy 361 -PMGPKLT-P-LFHSKYHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVVEFGQEH 417
Db 459 hi-ieetqlfdeqepvltlllskgrkfvyagsngvvqaplafeckhgt--cedcvlar 516
Qy 418 AFNIMEIOPFRRAAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDLCEVYGGCHGLMSR 475
Db 517 dpycawspptatcvahqtspgrglqemsdasvcpd-ks-kgsyrqhfkkhggaetael 574
Qy 476 DPYCGWDQ--GRCISYSSERSVLQSI--AEPHKECPNPKPKAPLQKVLAPNSRYL 532
Db 575 kcsqksnlarvfkfngvikaespyglmrknllif-nlsegdsgvycqlsae 628
Qy 533 SCPMESRHAYSWRHKENVEQSCEP--GHOSPNCILFIENLTAAQYGHYFCEAOE 585
RESULT 7
ID W51313 standard; Protein; 776 AA.
AC W51313;
DT 08-SEP-1998 (first entry)
DE Rat semaphorin W.
KW Rat; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW immunosuppressant; gene therapy; diagnosis; research reagent.
OS Rattus norvegicus.
PN W09815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMO) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-PSDB: V07279, V07280.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Claim 1; Page 60-64; 90pp; Japanese.
CC The present sequence represents rat semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 776 AA;

Query Match 11.6%; Score 518; DB 32; Length 776;

Best Local Similarity 27.0%; Pred. No. 5.28e-37;
Matches 156; Conservative 141; Mismatches 227; Indels 54; Gaps 38;

Db 43 rsvprtlpiseadstlrfaaashntynysallvdpashtlyvgardisfaltlpfsgerp 102

Qy 17 RSGPRI-FAVMKGVGDQDRVDFGQTPEHTVLFHEPSSVWVGGRGVY-L-FDFPECKN 73


```
QY 83 STKGSC-L-DKR--DCENVITLLRRSEG--LLACGTNARHPSCWN--LVNGT-VVPLG-E 133
Db 152 grglcfdpdmhstaiysegqlysatv-adfsqtdp-l1-yzg-p-lrtersdl-k-qln 204
QY 134 MRGYAPFSDENSLVLFEGDEYYSIRKOEYNGKIPFRPRIRGESELYTSDIVMNPQFI 193
Db 205 apfnv-tmeyndiffretaveincgkaly-srvarcvkhkdg-phgggdrtwtsf 261
QY 194 KATIVHQDAYDDKIYFFREDNPD-KNPEAPLNVSRVAQLCRGOGGESSLSVSKWNTF 252
Db 262 lksrincsvpgdpyfyneigdstiienyggq-vekllygvftpvnsggsavcfs 320
QY 253 LKAMLVCSDAATNK-NFNRLQDVF-LLPDP-SGQRDRTRVYGVFNPWN-Y--SACVYS 306
Db 321 mksilcsfdgpkgeqetmnsnwlavpslkvppepgqcvndrtlpdvsvnfkshtlmd 380
QY 307 L-G--D-IDKVPT--SSLKG-YHS--SL--PNPRGKCLPDQOPIPTTFQVADRHPEVA 356
Db 381 eavpafft-rpillrislqyrfkaiavdqvrtpdgkaydvlfigtdgdgkvikalnsaf 439
QY 357 QRVEPMGPKLTPLFH-S-KYHYQKAV-HRMOASHGETFHVLYLTDRGTIHKVVEPG-- 411
Db 440 dsdtdsvsviceqlvppgvklyvrmvmdgdsklvvssddelaiklhrcgskit 499
QY 412 EOEHSFAFNIME-IQPFRRAAAIQTM---SLDAERRKLYVSSQWEVSQVPLDLCVEYG-G 466
Db 500 nrecvslqdpvcawndvelkctavgsdws 530
QY 467 GCHGCLMSRDPVCGDQG-G-RCISYSSERS 495

RESULT 10
ID W51314 standard; Protein: 587 AA.
AC W51314:
DT 08-SEP-1998 (first entry)
DE Human semaphorin W.
KW Human; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PN Kikuchi K, Kimura T;
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PI (SUMU) SUMITOMO PHARM CO LTD.
PA Kikuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-PSDB: V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 9.6%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 2,20e-28;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgdelyffftetsrfdsyerikvrvarvcsgdlgrkrtlg-qrwtfklkadtllcp 115
QY 201 DOAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGOGGESSLSVSKWNTFLKAMLVCS 260
Db 116 gpehrasslqdvavrlpelga-gtlfyglfssqwgatavcfrpddirtvlnpg 174
QY 261 DAATKFNENRLQDVFLLPDPGQWRDRTRVYGVFNPWN---YSACVYSLGDIKVFRTS 317
Db 175 frelhdcnrglpvvdndvpqprgpcetnmnlrthfgsslsldrvlfridhplmdrp 234
```

```
QY 318 --SLK-GYHSSLP---N-----PRPGKCLPDQOPIP--TETFEQVADRHPE-V-AQR-VE-P 361
Db 235 vfpadgphllvttdtaylrvvahrvtslsgkeydvlylgtedghlhravrigaq-ls-vl 292
QY 362 MGPIUKT-PLFHSK-YHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVVEPGEHSEFAF 419
Db 293 edlal--fpeppqvenmkl-yhsw-llvgstrtevtqnttncgrlqs-csecilaqdpvc 347
QY 420 NIMEIQPFRRAAQTMSLDAERRKLYVSSQWEVSQVPLDLCVEYGGCHCLMSRDPYC 479
Db 348 awsrldcavahaghrglvqdiessadsvssicpkpegrpvrvfavataahvlpesps 407
QY 480 GWD-Q-GRCSIIYSERSVLQSIINPAEPHKPCPNPKDPKAPLQKVS LAPNSRYVLSCPME 537
Db 408 sawascvwhqpsgv 421
QY 538 SRHATYSWRHKENY 551

RESULT 11
ID W64221 standard; Protein: 974 AA.
AC W64221:
DT 06-OCT-1998 (first entry)
DE Human secreted protein from clone CUI45_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN WO9827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GEMY ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-362774/31.
DR N-PSDB: V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CUI45_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 8.7%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 8.73e-25;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechnfikvlknddalfvcgtnafnpncrnykmtlepfqdefsgmarcpdyakhan 172
QY 91 KRDCENIT-LLERRSGLACGTNARHPSCWNLVNGTVVPLG-EMRGYA--PSPDEN 146
Db 173 valfadklysatvt-dflaidavlyrsl-gesp--tlrtvkhdskwlkpeyfvavdyg 228
QY 147 LVLPFEGDEVSTIRKQEVNG-KIPFRIRIGESELYTSDTVMQNQPFIKATIVHQDAYD 205
Db 229 dylyffirelaventyngkvfvrvaqckndmgsgsrviekwtsfkarlncsvpgds 288
QY 206 DKIIYFFREDNPDKNPEAPLNVSRVAQLCRGOGGESSLSVSKWNTFLKAMLVCSDAATN 265
```


QY 367 -TFLFH--SKYHOKYAVHMOASHGETFHVLYLTTDRGTIHKVVERPEQEHSEAFNIME 423

Db 470 eIdaysparcsgkrtactarriIqleIdteghrIlfafsgclvyIpliscarhga-cqrs 528

QY 424 -IQPF---R---RAA--A--IOTMSLDAERRKLYVSSOWEVSQVPLDLCEVYGGGCH-G 470

Db 529 clasgddpycgwhsrgcvtlrgsggtcdvdqagngesmehdcdq 572

QY 471 CLMSRDPYCGWDOGR-CISIYSSE-RSVLOSINPAE-PHKECPN 511

Search completed: Thu Jul 8 19:13:42 1999

Job time : 39 secs.

THIS PAGE BLANK (USPTO)

QY	KTPLFHSKHYHQA	VAHRMQASG	GEFPH-VL	YLITDRGTH	HTKHYVEG	EDG	SHAF	INMI	424																																										
Dp	479	NPOKAPAI	IONIL	LDNTNL	KLYNSE	WESE	VEPL	DLCS	YVND	CSEMSR	DLCT	WYNN	538																																						
QY	425	QPFRR	AAIQ	TMS	LDAR	RRLY	VSQ	WEVE	SQV	PLD	CEYV	GGG	CHGLMS	RDPY	CGMDG	484																																			
Dp	539	TC-S-FK	-ORS	VS	TGG	PAR	RTL	SEMG	CDH	YAP	TVY	KHQS	ITPL	NS	LS	CPAV	SNAD	595																																	
QY	485	RCIT	YSSEK	VL	OS	IN	PA	EPH-KE	-CP	NK	PKAP	LAQ	CS	LA	PN	RY	LS	CME	SRAT	542																															
Dp	596	YFMT	KDGF	TEK	RCH	VK	TKH	KD	CL	LL	AN	ST	AT	NG	TH	VC	NK	EDS	640																																
QY	543	YSMH	KE	NE	VS	CE	PH	QSP	NC	IL	FI	EN	LA	QY	GH	FC	BA	OG	CS	587																															
RESULT	2																																																		
ENTRY																																																			
TITLE																																																			
ORGANISM																																																			
DATE																																																			
ACCESSIONS																																																			
REFERENCE																																																			
#authors																																																			
#journal																																																			
#title																																																			
#cross-references																																																			
#accession																																																			
#status																																																			
##molecule_type																																																			
##residues																																																			
##cross-references																																																			
GENETICS																																																			
CLASSIFICATION																																																			
SUMMARY																																																			
Query Match																																																			
Best Local Similarity																																																			
Matches																																																			
Dp	112	ECMNFV	LLAHV	NH	THLL	AC	RTGA	FH	PTCA	LM	FW	AT	AG	TH	AS	TG	PE	KLE	DG	KTP	PPD	171																													
QY	93	DCENV	IT	IL	LR	RR	SSG	-LL	AC	GT	NA	RH	PS	C---	WNL	VNG	I-V	-V	-P	-L	GE	RG	YAP	ESP	142																										
Dp	172	RHRP	SV	LV	GE	LE	ISG	-T	AD	LG	RO	DT	IF	RS	I	GO	PD	SR	TR	EP	HD	SR	ML	NE	PK	FV	FI	230																							
QY	143	DENV	LV	FE	GE	DE	VY	STR	KOE	Y	NGK	-I	P	FR	R	T	R	G	E	S	E	L	T	-S	-D	T	V	-M	ON	O	F	I	A	T	-I	197															
Dp	231	PESE	N	P	D	D	K	I	E	F	F	R	E	S	A	V	E	A	P	A	P	A	R	G	M	S	V	R	Q	I	C	R	N	D	L	G	O	R	S	L	-V	N	K	T	T	F	L	K	A	R	289
QY	198	VH	Q	O	A	V	D	D	K	I	E	F	F	R	E	D	N	P	K	N	P	E	-A																												

Qy	471	CLMRDPRCGMDGRCRSTISYSSESTVL--QSINPAREHKKSCRPKRDKAPLQKVLAPLAPS	528
Db	585	GSAPLECEPNSLQAHVOM	602
Qy	529	RYLLSCPMESRHAATYSM	545
RESULT	3		
ENTRY	148748	#type complete	
TITLE	semaphorin E - mouse		
ORGANISM	#format_name Mus musculus #common_name house mouse		
DATE	02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 04-Sep-1998		
ACCESSIONS	148748		
REFERENCE	148744		
#authors	Puschel, A.W.; Adams, R.H.; Betz, H.		
#journal	Neuron (1995) 14:941-948		
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension		
#cross-references	MUID:95267431		
#accession	148748		
#status	preliminary; translated from GB/EMBL/DBJ		
#molecule_type	mRNA		
#residues	1-751 #label RES		
#cross-references	EMBL:X85994; NID:g854331; PID:g854332		
GENETICS			
#gene	sem		
CLASSIFICATION	superfamily semaphorin		
SUMMARY	#length 751 #molecular-weight 85259 #checksum 8961		
Query Match	16.3%; Score 725; DB 2; Length 751;		
Best Local Similarity	33.0%; Pred. No. 5,98e-132;		
Matches	148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;		
Db	167	SENNVNTVSYMTIEELFSGMYT-DEMGTDAAFRSLTKRMOLRTDOHNSKWLSEPMFVD	225
Qy	139	PFSDENSLVLFEGDEYVSTIRKQEVNGK-IIPFRFRIRGESELYTS--DTV-MONQOFIK	194
Db	226	AHVIPDGDIPDAVYVYFFEKERLTDNNRSKQIHSMIARICPMDTGORSL-VNKKWTFEL	284
Qy	195	ATIHQD-QAADKILYFFREDNDPKRPEAPLAVSRVAQLCRDGCGESSLSKKNTEFL	253
Db	285	KARLVGSVTDDEGPEHFDELEDFLL-ETDNP-RTLLVYGIFTTSSSVFQSGAVCYHL	342
Qy	254	KAMLVCS--DA-ATNKNFNRLQVLEFLPDSSGQMPRQTVRYGVFSNWN-Y--SACVYSL	307
Db	343	SDIOVTEGPAHKEGNNHQLISQGIPIPRPGTCGGAFTPMNMETTDEPDDVYTFIR	402
Qy	308	GDIDKIVRTS-SLK-G-----YHSLPPIPRGKCLIPDO-QP-I-PTETF--QVADRRP	353
Db	403	NHPLMYSISPRIHRRLIVRIGTDYKTKTAAVDVNAADG-RYHVLFLGDRGTQGVVV	461
Qy	354	EVAQVRPEMGL-KTPIF-H--SKYHQKVAVHRMAQSHGETFHVLLITTDRETIIKVE	409
Db	462	LPTSSASGELLLELEVFKNHPYITTMELSSKKQOLYSSNGCQVSYSLRCHITGTGC	521
Qy	410	-PGQHSFSAFNIMEIOPFERRAALIQTMSIDAERRKLYVSSQWVEQVPLDCELVYGGCC	468
Db	522	ADCCIADPCAMGDSHGSRYFPTGKRSRROVNRGNPLTQCRGNELKAYRNAAEITVOY	581
Qy	469	HGCLMSADPYCGMQMGRCISITYSSE--RSVLQSLINAEHPHKECP--NPRPKARPLQKVS	524
Db	582	GVRNNSFFLECAPKSPQASIKWLQDK	609
Qy	525	A-PNSRYLLSCPMESRHAATYSM-RKEN	550
RESULT	4		
ENTRY	G01856	#type complete	
TITLE	semaphorin V - human		
ORGANISM	#format_name Homo sapiens #common_name man		
DATE	21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 04-Sep-1998		

Accessions	G01856	G08634	Sekido, Y.	Submitted to the EMBL Data Library, June 1995
REFERENCE				
#authors				
#submission				
#accession	G01856			
#status				
#molecule	mRNA			
#residues	1-749	#label	SRK	
#cross-references	EMBL:028369	NID:	g9742283	PID:g9742284
CLASSIFICATION		#superfamily	semaphorin	
SUMMARY		#length	749	#molecular-weight
			83121	#checksum
				2747
Query Match	16.1%	Score	716	DB 2: Length
Best Local Similarity	33.0%	Pred	No. 7,10e-130	
Matches	167	Conservative	115	Mismatches
			170	Gaps
				41
Db	112	ECMFNFKLLAHYNTHTLACGTGAFHPTCAFEVGHRAEFPVLTLDGRIEDGKSPYD	171	
Qy	93	DCENTITLLER-RSEGLIACGTNRHSPC-WNLV-N-G--TVVPL--G--EM-RGAPFS	141	
Db	172	PRHRAASVVEBELYGYA-ADLGRDFTTIRSGQPSLTPEPHDSRWLNEXFVAVFW	230	
Qy	142	PDENSLVFEDEDEVYSTIRKOEYNGK-IPRRRIRGESELYT-S-DIV-MONPOFIKAT-	196	
Db	231	IPSENPDDDKTYFEFRRTAVEAPALGRLSVSVGOICRNDVGGQSL-VNKKTFELKA	269	
Qy	197	IVHQQAIDDKTYFFREDNDPKNE-APLWASVVAOLCRDQGGESSLSYSKWNTEKA	255	
Db	290	RLVCSVPVEEDTHEDQADVFELS-SHDH-RTPLLYAVFTSSSIFQGSAYCVYKNDV	347	
Qy	256	MLVCS--DAATNKNNRRLQDVFLLPDPSCGQRDRIRVGVENPNKN-Y--SAYCVYSIGDI	310	
Db	348	RRAFLGPAHKEGPHQWVSYQGRVYPRPGMCDSTFTGTFSSTKDEPDDVIOAFRNHPL	407	
Qy	311	DK-V-F--RTSSL-K--GYHSSLPNRPCKCLPDQ-QPIP-TETP--GVAD--R-HPE	354	
Db	408	MYNSVLPFGQ-R-PLFLQGANITYFTQIAAARVAADAQH-IDVLFITDGTVLAKYSVP	464	
Qy	355	VAQREPEPGRPLKTPLE-H--SKYHQAIVARMQASHGETHVLYLTLDRTHTIKVVE-P	410	
Db	465	KGSRPSAGCLLLEELHVEDSAAVYSWQISSKRRLQVYASASVAQIATLHNCAGHRCVT	524	
Qy	411	-GEOSHSAFNINMETOPRRRAAIOTMSLDMEKKRLVSSWMEVSQVPLDCEVYGGGCH	469	
Db	535	ECCLARDPYCAMDVACTRFPQSAKRREPRDVRNGRDPSTLCSGSSRRPALLEHKEVGE	584	
Qy	470	GCLMSRDYCGMDQRCISIVSSERSVL--QSINPAEDHKKCPNPKDKAPL-QKVS LAP	526	
Db	585	GSAFLCECPRSLOARVWTFORACV	610	
Qy	527	NSRYLLSCPMSRHAIVSMR-HKENV	551	
RESULT	5			
ENTRY	G02173	#type	complete	
TITLE			semaphorin III family homolog - human	
ORGANISM			formal_name Homo sapiens #common_name man	
DATE			21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change	
			04-Sep-1998	
ACCESSIONS	G02173			
REFERENCE	G09275			
#authors	Naylor, S.			
#submission				
#accession	G02173			
#status				
#molecule	mRNA			
#residues	1-753	#label	NAV	
#cross-references	EMBL:038276	NID:	g1061350	PID:g1061351
CLASSIFICATION		#superfamily	semaphorin	
SUMMARY		#length	753	#molecular-weight
			84941	#checksum
				5681
Query Match	15.4%	Score	688	DB 2: Length
Best Local Similarity	31.4%	Pred	No. 1.97e-133	

Matches	39:	Conservative	121:	Mismatches	140:	Indels	42:	Gaps	30:
Db	171	PYDKLDTASLINDELIAGYIT-DEFGTDAIRITLQKQAMTDOYNSRWLNDSPEFIH	229						
Qy	139	PFSPSPENLFLFDEDEYVSTIRKOEYNGK-IPRRRIRGESELYTS--DTV-MONPOFIK	194						
Db	230	AEIIPDSAMENDKLYFFERRSAAE-AQSPAVVARIIRICLNDGGHCC-LVNWSPFLK	287						
Qy	195	ATVHQDAVDKITYEFREDNPKNPAPLANSVAOLCGDGGSSLSVSKWTFELK	254						
Db	288	ARLVCSYVGEDGIETHDELQDEV-OOTODV-BNPIYAAFTSSGVFRSACVYSMA	345						
Qy	255	AMLVCS---DAATKNFNRLODVFLPDPSCQWMDTRVYGVFSNP-WNY--SACVYSLSG	308						
Db	346	DIRVYNGPFAHKCGPNYQWMPESGKMPYPRPGTCPCGTFTPSKSTKDYIPEVINFRS	405						
Qy	309	DIDKVFRTS-SLK-G-YH-----SS-LPNPRPGKCLDPDQ-OP-I-PIETF--GVAD--R	351						
Db	406	HPLMQAYPIQORRPLVVRT-GARYRLTTIADVQDSADG-RYEVLLGDRGVQKVIY	463						
Qy	352	HPVAAQREPEMG--PLKTPLFHSYHYOKAVAHMQASHGTFHVLITTDRTGTHHVE	409						
Db	464	LPKDOEWEEMLEEEVEFKPAPVKTITSSKROOLYVASAVTLLSHRCQAYAAC	523						
Qy	410	-PGEDEHFAFNINMIEIOFRRMAALIQMSLDAERKLYLVSSQWESQVPLDCEVYGGC	468						
Db	524	ADCCLDARDPYCAMDQACSRYSATSSKRRSRQDVRHGNPLRQCRGFSNANKNAVESQY	583						
Qy	469	HGCLMSRDPYCGMDQGCIS-IYSESR-SVLQSNINPAPHEKCP--NPKPDKAPLQKVL	524						
Db	584	GVAGSAALFECQPRSPQATVKW	605						
Qy	525	A-PNSRYLLSCPMESRRHATYSM	545						
RESULT	6								
ENTRY	A49069	#type complete							
TITLE	collapsin - chicken								
ORGANISM	#formal_name Gallus gallus #common_name chicken								
DATE	07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 04-Sep-1998								
ACCESSIONS	A49069								
REFERENCE	A49069								
#authors	Luo, Y.-; Raible, D.; Raper, J.A.								
#journal	Cell (1993) 75:217-227								
#title	collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones.								
#accession	A49069								
#status	preliminary; not compared with conceptual translation								
#molecule_type	mRNA								
#residues	1-772 #label LUO								
#cross-references	GB:U0528; NID:9410078; PID:9410079								
CLASSIFICATION	#superfamily semaphorin								
SUMMARY	#length 772 #molecular_weight 88867 #checksum 9712								
Query Match	15.3%; Score 681; DB 2; Length 772;								
Best Local Similarity	31.5%; Pred. No. 7,96e-122;								
Matches	146; Conservative 119; Mismatches 150; Indels 48; Gaps 38								
Db	166	RKGSYDPKLLTASLINDGELYSGTA-ADENGRDPAIFRTLGHHNPIRTEOHDSRWLNDP	224						
Qy	135	RGYAFSPDENSLVLFEDDEVYSTIRKOEYNGK-IPRRRIRGESELYTS--DTV-MQNP	190						
Db	225	REISAHLPESDNPEDDKIYFFPEFEMALDGHNTKATHARIGOLCKNDGFRSL-VNWK	283						
Qy	191	QPIKTIYHQ--DQADADLIYFFREDNDPKPEAPLANSVAOLCRGGQGGSSLSVSKW	249						
Db	284	TTFLKARLIGSVPGNGIDTFHDELQDEVFLK-NSKDP-KNPYVGVFTTSSNIFKSGAVC	341						
Qy	250	NTEFLKAMLVCS--DA-ATNKNFNRLODVFLPDPSCQWMDTRVYGVFSNPWN-Y--SAYC	303						
Db	342	MYSMTDVARVELGPAHRDGPNYQGVYPRPGTCPSKTFGGFGDSTKDLDEVIY	401						

Oy	304	VYSLGDDIDKIVRTS-SLK-G--YH--S--S-LPNRPCKLPDO-QPl-PIETP--QV-A	349
Db	402	FARSHAPMNPFPINS-RPIMIKTDVDYQFTQIVVDVDAEDGO-XDVMFIGTDIGTVL	459
Oy	350	-DR-HPEVAQRVERPGPLKTPLEHSHK-YHYQKAVAHMQASHGETFHVLYLTDRGTH	405
Db	460	KVYSIPKRTWHELEVLLEEMTVFREPPIVISAUKISTQOOLYIGSATGVSQPLHRCDV	519
Oy	406	KVVE-PGEOHSEAFNINE-IQPFRAAAIOTMSLDAERRLYVSSQWESQVPLDCEV	463
Db	520	YKACAECCLLARDPYCAMDSSCSRFPFTAKRRTRRODIRNGDPLTHCSDDQHHNDSSGQ	579
Oy	464	YCGGCHGLMSRDPICGMDGRCISYISSE-RSVL-QSINAPAEHKECPN-PKFDKAP--	518
Db	560	TLEEKIYGVENSSTFLECSPEKSORAIYV-WQFOKONDHKEV	621
Oy	519	LQ-KVSLA-PNSRYLLSC-PMESHATYSNR-HKEVNEQSC	556
RESULT	7		
ENTRY	158169	#type fragment	
TITLE	semaphorin III - mouse (fragment)		
ORGANISM	#formal_name Mus musculus #common_name house mouse		
DATE	26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 04-Sep-1998		
ACCESSIONS	158169		
REFERENCE	158169		
#authors	Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.; Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.		
#journal	Neuron (1995) 14:949-959		
#title	Semaphorin III can function as a selective chemorepellent to pattern sensory projections in the spinal cord.		
#cross-references	MUID:95267432		
#accession	158169		
#molecule	#stictus preliminary; translated from GB/EMBL/DBJ		
#residues	1-666 #label RBS		
#cross-references	GB:L40484; NID:9703189; PID:9703190		
GENETICS			
#gene	SemaIII		
CLASSIFICATION	#superfamily semaphorin		
SUMMARY	#length 666 #checksum 9654		
Query Match	15.0%; Score 667; DB 2; Length 666;		
Best Local Similarity	30.1%; Pred. No. 1.30e-118;		
Matches 150; Conservative 130; Mismatches 168; Indels 50; Gaps 39;			
Db	60	RGKSPYDPKLTASLIDGELYSGRA-ANFNGRQFAIFRTIGHHNPIRTEDHDSRWLNDP	118
Oy	135	RGYAFSPSDENSLVLEGEDEVYSTIRKQOYNGK-IIPRRIRIGSESLYTS-DIV-WQNP	190
Db	119	RFSIAHLIPESDNPEDKVVYFFPRENAIDGESHKATGARIGQICKNDFGGHRSL-WNKW	177
Oy	191	QIKATIVHQ-DQAYDKQIYFFREDNDNKRPEARPLANSRAQJCLCRDGGESSLSYSKM	249
Db	178	TFPLKARLIGSPVPGNGIDTHFDELQDVFILW-NSKDP-KNDIYVGVTTSNIFKSGAVC	235
Oy	250	NTEFLKAMLVLS-DA-ATNKNFNRLQDVFLLPDSGGQKRDTRYGVSPNPN-Y-SAVC	303
Db	236	MYSMSDVARVLGLPAAHHDGPNYQGNVPYPRQCTCSKTFGGGDSKLDLPDVIY	295
Oy	304	VYSLGDDIDKIV--F--RTSS-LK-GYHSELPLNRPCKLPDO-OPi-PIETP--QV-A	349
Db	296	FARSHAPMNPFPINN-RPIMIKTDVYQFTQIVVDVDAEDGO-XDVMFIGTDIGTVL	353
Oy	350	-DR-HPEVAQRVERPGPLKTPLEHSHK-YHYQKAVAHMQASHGETFHVLYLTDRGTH	405
Db	354	KVYSIPKRTWHELEVLLEEMTVFREPPIVISAUKISTQOOLYIGSATGVSQPLHRCDI	413
Oy	406	KVVE-PGEOHSEAFNINE-IQPFRAAAIOTMSLDAERRLYVSSQWESQVPLDCEV	463
Db	414	YKACAECCLLARDPYCAMDSSCSRFPFTAKRRTRRODIRNGDPLTHCSDDQHHNDHGP	473

Oy	464	YGGCGHCLMSRDYPICMWDGRCISIIYSSE-RSVL-QSINPAEPHKECPN--KPKDK--AP	518
Db	474	SLEERIIAYGVENSSTFPECSPKSORALVYQFORNEDRKKEIKMGDHIIRTEOGILLRS	533
Oy	519	LQ-KVSLA-PNSMYIYLSG--PMESRHATY-SW-RHKENVEQSGEPE-HQ-SPNCLFLFEN	570
Db	534	LQKDSGNVILCHAVEGFG	551
Oy	571	LTAQOYGHYFCEAGEGSGY	588
RESULT	8		
ENTRY	148747	#type complete	
TITLE	semaphorin D - mouse		
ORGANISM	#format_name Mus musculus	#common_name house mouse	
DATE	02-Jul-1996	#sequence_revision 02-Jul-1996	#text_change 04-Sep-1998
ACCESSIONS	148747		
REFERENCE	148744		
#authors	Puschel, A.W.; Adams, R.H.; Betz, H.		
#journal	Neuron (1995) 14:941-948		
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension		
#cross-references	pubmed:95267431		
#accession	148747		
#status	preliminary; translated from GB/EMBL/DBJ		
#molecule_type	mRNA		
#residues	1-772	#label RES	
#cross-references	EMBL:X85993; NID:g854329; PID:g854330		
GENETICS			
#gene	semd		
CLASSIFICATION	#superfamily semaphorin		
SUMMARY	#length 772	#molecular_weight 88710	#checksum 1776
Query Match	15.0%;	Score 669;	DB 2; Length 772;
Best Local Similarity	30.28;	Pred. No.4.51e-119;	
Matches 170;	Conservative 143;	Mismatches 182;	Indels 67; Gaps 51;
Db	58	HTFLDLE-ERSRLVVGAKDHFISFNVLNIDFQKIYWPVSY--TRDECKWAGKDIKCECA	115
Oy	43	HTVLFHERGSSVWVGGRKVIYLFDEPDEGN-AS-VRYINISTGKSGC--LDK---RCE	95
Db	116	NFIVLEAIVQTHLYACGTAHFPICTYLEVGHHPEDNIFKLQDSHFENGKSPYDKL	175
Oy	96	NYITLLERRSBG--LLACGTNARHPSCMNL-V--N--GTVVPL--G--EK-RGYAPFSPDE	144
Db	176	LTAULLIDGELYSTA-ADFMGDPAIFRLGDHNRIRFQDHSKMLNDPRITSHLPE	234
Oy	145	NSLVLEFGDEYVSTIRKQEVNGK-IPRFRIIRSESLYS--DIY-MONPQIKATIYHQ	200
Db	235	SDNPEDDKYVFFFEFNALIGEHSGKATHARIGQICKNDGGRSL-VNKTWTFILKARLIC	293
Oy	201	-DQAYDKIITYFFREDNPDKNPAPLANSVVALCGDGGGSSLSYSKWNITFLKAMLYC	259
Db	294	SVPGPNIDITHFDELQDVFIL-NSKDP-KNPIYGVFTTSSNIFKGSAYCMYSMSDVARY	351
Oy	260	S--DA-ATNNKNRNLDQVFLPDPGSGQMDTRYGVFNSPMN-Y--SAYCVYSLGDDIKV	313
Db	352	FLGAYARDGPNQWVPYQGRVYVPRPGTCPSKTEGFGFSTQDLPDDYITGRSHPMYIN	411
Oy	314	FRTS-SLK-G--YH--S--S-LPNPRPGKCLPDQ--QPI-PTETF--QV-A-DR-HPEVAQ	357
Db	412	PVPEINN-RPIIMIKTDVNVQFTIYVDVRYDAEDGQ-YDVMFICGTVGVYGLKVSYPKRETW	469
Oy	358	RVEPMGGLKPLPLHS--KTHYQVAAVHRAQASHGEFHHLYLTTRGTTHAKVE--PGEOE	414
Db	470	HDLEEVLEEMTYFRRETTISAMELSTKQOQLYIGSTAGVADLPRLHRDIIYKACACECL	529
Oy	415	HSFAFNIME-IQPFRRAAAIQOTMSLDAERKLIVSSQWVSGVPLDLCEVYGGCGHGLM	473
Db	530	ARDPYCAMDSSGSRFFPIAKRTTRQDIRNGDPLTHGSDLEDHNDHHPGLEERIIYGV	589
Oy	474	SRDPYCMWDGRCISIIYSSE-RSVL-QSINPAEPHKECPN--KPKDK--AP-LQ-KVSLA-	525

ENTRY	9	#type complete
530	ENSSTFLECSPKSORALYV-MQ 610	
526	PNRSRYLSC-PMESRHATYSMR 546	
RESULT	9	
ENTRY	D49423	#type complete
TITLE	semaphorin III precursor - human	
ORGANISM	#formal_name Homo sapiens #common_name man	
DATE	06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 04-Sep-1998	
ACCESSIONS	D49423	
REFERENCE	A49423	
#authors	Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.	
#journal	Cell (1993) 75:1389-1399	
#title	The Semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules.	
#accession	D49423	
#status	preliminary; nucleic acid sequence not shown	
#molecule_type	mRNA	
#residues	1-771 #label KOL	
#cross-references	GB:L26081; NID:g799328; PID:g436560	
GENETICS		
#gene	GDB:SEMA1	
CLASSIFICATION	#cross-references GDB:283448	
SUMMARY	#superfamily semaphorin #length 771 #molecular_weight 88889 #checksum 6249	
Query Match	14.8%; Score 661; DB 2; Length 771;	
Best Local Similarity	30.3%; Pred. No. 3,07e-117;	
Matches	151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;	
Db	166 RGSHPDPLKLTASLIDGELYSGA-ADFMGRDFAIFRTLGHHNPRTEDHDSRWINDP 224	
Qy	135 RGYAFSPDENSELVLEEDVEYSTRKOEYIGK-IPRRRRRGSELYTS--DIV-MQNP 190	
Db	225 KFI SAHLIESDNPEDEVYFFEFRENAIDGESHKATHARIGQICKNDFGGRSL-VNKK 283	
Qy	191 QEIKATIVHQ-DQAVDDKIIYFFREDNDKKNPEAPLVNSRYAQCRRDGGESLSLWK 249	
Db	284 TTFLEALRICSPGNGIDTFDELQDYFLM-NKRD-KNPVYGVTTSSNIFKGSAYC 341	
Qy	230 MTFLEALRYCS-DA-ATNKNFNRLQDYFLLPDSGGQRDTRVYGVFSNPNV-Y--SAYC 303	
Db	342 MYSMSDVRVFLGPAHHDDGPNYQGVNRPVPRGTQSKTFGGFSDKQDPRDVIIT 401	
Qy	304 VYSTLDDIKVTRTS-SLK-G--YH-S--S-LPNRPREKCLPQD-QPI-PETF--OV-A 349	
Db	402 FARSHAPMYNPVFPNNRNPIIKTDVNYQFOTIIVADRYADEGQ-YUMFEGTDVGVTLK 460	
Qy	350 -DR-HPEVAQQRERPGRLKPL-FHSKHYKQKVAVHMQASHGTFHVLTLTDRIQTHK 406	
Db	461 VYSTKEWYDLEEVLEEMTVFREPRASISAMELSTQAOOLYIGSTAGVADQLRHRCDIY 520	
Qy	407 VVE-PGEDEHSAFANIME-IQPFRAAAIQTMSDAERKRLYVSSQEVESVQPLDCEVY 464	
Db	521 GKACAECCLARPYCAMGSCASRYFPAKRRTRQDRIANDPLTLHOSDLHNDHGHSP 580	
Qy	465 GGGCGGCLMSADPTFCGMDOGRCIYSSE-RSYL-QSINPAERPKECRPNKPRD---AP 518	
Db	581 EERITVGENSTPFLCSPKSQORALVYWOFORNREERKEELIRVDNHIIRTDGILLRSLO 640	
Qy	519 LQKVASLA-PNSRYYLSC-PMESRHATY-SM-RHKEAV-EGGCEGEGHQ-SPNCILFIENLT 572	
Db	641 QKDSNGNYLCHAVEHGTQ 658	
Qy	573 AAOYGHYFCEAOEGSYFR 590	
RESULT	10	
ENTRY	I48746	#type fragment
TITLE	semaphorin C - mouse (fragment)	

ORGANISM	#format_name Mus musculus #common_name house mouse			
DATE	02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 28-Feb-1997			
ACCESSIONS	I48746			
REFERENCE	Puschke, A.W.; Adams, R.H.; Betz, H.			
#authors	Neuron (1995) 14:941-948			
#journal	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension			
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension			
#cross-references	MIMD:95267431			
#accession	I48746			
#status	Preliminary: translated from GB/EMBL/DBJ			
#molecule_type	mRNA			
#residues	1-782 ##label RES			
#cross-references	EMBL:X83592; NID:9854327; PID:9854328			
GENETICS				
#gene	semc			
SUMMARY	#length 782	#checksum 1571		
Query Match	13.9%	Score 618:	DB 2:	Length 782:
Best Local Similarity	33.7%	Pred.	M.2.04e-107:	
Matches 149:	Conservative	96:	Mismatches 148:	Indels 49: Gaps 32:
Db	74 KRDCQNYKILPLPNSHSLTGCFAAFSPICAYIHASFLLADENAGVILDEGKHCF 133			
Qy	91 KRDCENITLTLERSSEGLACGTNARHPC-W-N-----LV-NGIV-VPLDEMGIYAF 140			
Db	134 DPNFKTALVYDGLYGTIVS-SFGQNDPAISRQSSRPTKTESSLNMLDPAFVASAT 192			
Qy	141 SPDENSLVLFEGDEYVSTIRKQENKIPFRIRIGSELSYSDIV--MQNPQIFKATIV 198			
Db	193 PESLUSPGDDDKYIFFFSETGQEFEEFNTI-VSNVARYCKGDBGGERVLO-QKWTSL 250			
Qy	199 HQDQ-A-Y--DDKYFFREDNDP-KNPEAPLVNVAALDGGDGGESSLSVSKMFTL 253			
Db	251 KAOLICSRPDGPFENVLODFLLNPQQ-DWKRKTLISGVFSSQMRGRTGESSAICVFTM 309			
Qy	254 KAMLYCSDAATNKKFNKLQDFLL-DPSSQMDITVYGVFSIPMN--Y--SAVCVSL 307			
Db	310 NDVQKAFDGLKKKNRFTQOMYETHQVPRPGACITNSAREKRINSLOLDPVLNFI 369			
Qy	308 GDIDKVF-----RRSS-LKG-Y-HSS-LPRPRGKCLPD--QQPIPTETFFVADAHPEVA 356			
Db	370 KDHFLLMDQVARSRLLLQPRARQVAARVAPGLHS-TYDVLFLGTGGRLKAVATLSSR 428			
Qy	357 ORVEPM-GPLETPLF--HSKYHYQKAVAHMQASGHETFHVLVLTTRDRIHKVPEQEQ 413			
Db	429 VH---I-IEELQIPQGPQVQNLILDSHGGLYASHSQVQVPAVANCSLYPT-CGDDLL 483			
Qy	414 EHSFAFNMEIQPRRAAIIQMSLDAERRKILVSSQMSQVPPDLCEVYGGCGHGLM 473			
Db	484 ARDPYCAWTSACRLASTLYOPD 505			
Qy	474 SRDPYCGMDGRC-I-SIYSSE 493			
RESULT	11			
ENTRY	S66498 #type complete			
TITLE	M-sema F protein precursor - mouse			
ORGANISM	#format_name Mus musculus #common_name house mouse			
DATE	28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 10-Sep-1997			
ACCESSIONS	S66498			
REFERENCE	S66498			
#authors	Inagaki, S.; Furuyama, T.; Iwahashi, Y.			
#journal	FEBS Lett. (1995) 370:269-272			
#title	Identification of a member of mouse semaphorin family.			
#cross-references	MIMD:95385809			
#accession	S66498			
#status	Preliminary			
#molecule_type	mRNA			
#residues	1-834 ##label INA			
#cross-references	EMBL:S79463; NID:g1110598; PID:g1110599			

FEATURE
1-21 #domain signal sequence #status predicted #label SIG
22-834 #product M-sema F protein #status predicted #label MAT
SUMMARY #length 834 #molecular-weight 92556 #checksum 7189

Query Match 11.7%; Score 522; DB 2; Length 834;
Best Local Similarity 28.5%; Pred. No. 1,06e-85;
Matches 142; Conservative 118; Mismatches 190; Indels 49; Gaps 36;

Db 102 KGKS-NQECFNFIRFLQPVNSSLVLCSTVAFQPKCTYINNLTFLDAFEEDGKCP 160
QY 85 KSCSDRCRCENYITLLER-RSEGLACCTNARHPSG-W-NLVNGVIV-VPLGEMGVAP 139
Db 161 YDPAGKHGGLVDGELYSATLN-NFLGTEPVILRYKWHSHSIXTEYLAFWLNPHVGS 219
QY 140 FSPDENSLVLFEGDEYVSTIRKQEVNGKIPRRIRIGSESELYTSD--TV-MONPQIKAT 196
Db 220 FYPEVSGFTGDDKITYFFSERAVEYDCYEQVAVARVCKGDMGAKRTQ-KKNTTF 278
QY 197 IVHQP-QAY--DDKITYFFREDNPDKNEAPLNSRVVQLCRDGGGSSLSVSKWTF 252
Db 279 LKARLVCSAPDKVYFNQKAVHTLRGAS--WHNTTFEGVFOARMQMDLSAVCEYQLEO 336
QY 233 LKAMLYCSDAATANKNFNRQADYFLRPSGQHRDTRVYGVFSNPW--NYSACVYSLGD 309
Db 337 IQOVEGPKYKEYSEQAOKWARYTDPVSPRPSGCIINMHDRNGYTSLELDPNTLFIKK 396
QY 310 IDKVF---RT-S--SLK-G-YHSSLPNRPCKLPD-QQPIP-TETGVADRHPV-VA- 356
Db 397 HELMEDQYKPRGRPLVLYKNTNFHYVADRPGLDGATYTYLFTIGTGMLKAVSLCP 456
QY 357 QR-VE-PMGP-LKTPLEFHSKY-HYOKVAVHRMQASHGEHFHVLTYLTDGTIHKVPEE 412
Db 457 WTH--M-VEELVQDOEP-VESELVSGSKVLFAGSRQOLQSLADCTKTRF-CVDCV 510
QY 413 QHSHAFNFMETQPRRAAIOTMSLDARRLKLYVSSQWEQVPLDCEVYGGGCHGL 472
Db 511 LARDPYCAMNVTNRCAVATTSGRSGSFLVQVHANTDSKMCNOYGIKVRISIPKNTVVS 570
QY 473 MSRDYCGMD-Q-CRCISITYSER-SVL-QSINPAEPHKECPNPKDKA-PLQK-VSLAP 526
Db 571 GTDYLPLCHLSNLAAHW 589
QY 527 NSRYLSCPMESRHAATYSW 545

RESULT 12
ENTRY 148745 #type complete
TITLE semaphorin B - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 28-Feb-1997
ACCESSIONS 148745
REFERENCE 148744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession 148745
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-760 #label RES
#cross-references EMBL:X85991; NID:g854325; PID:g854326
GENETICS
#gene
SUMMARY #emb #length 760 #molecular-weight 83458 #checksum 2188
Query Match 10.8%; Score 480; DB 2; Length 760;
Best Local Similarity 29.2%; Pred. No. 2.56e-76;
Matches 157; Conservative 121; Mismatches 203; Indels 57; Gaps 41;
Db 123 OCFNIRVLVSNATHLACGTFAPSPACTFIELDSSLPLILDKVMDGKQSPILTLFT 182

QY 93 DCENITLLER-RSEGLACCTNARHPSG-W-NLVNGVIV-LGE-W-RGAPFSP-D 143
Db 163 STQAVLVNG-MLYSGT-NNNFGLSEPIIMRTLGSHPVLTQIFIRMLH-ADASFVAALPS 239
QY 144 ENSLVLFEGDEYVSTIRKQEVNGKIPRRIRIGSESELYTSDTWNQNPQIFATIVHDDA 203
Db 240 -TQVVYFFFEETASBDFEELLYISRVAVQCKNDVGEKLLQ-KKWTFLKAOQLCAOPG 297
QY 204 YDCKIYFFREDNPDKNEAPLNSRVVQLCRDGGGSSLSVSKWTFELKAMLVCSDA 263
Db 298 -QLPNTIRHAILPADS-P-SVSRIVAFVTSQWQVGTSSASACASLTIDERYFKY 354
QY 264 TNKNRRLQADVFLRDPGQHRDTRVYGVFSNPWY----SAVCVYSLGIDIVFR-T- 316
Db 355 KELNKTSMWTYRGEVSPRPGSCSMGSPSDKALTF-MKD-HFLMDHVVGT-PL--TL 408
QY 317 SSL-K--G----YHSSLNPRPKCLPDQPIPTETGVADRHPVAVORVPRPKLTP 369
Db 409 VKSGVETRLAVESARGLDGSSHVVMYLGTSGLPKHAYV--QDSS-AVLVEEIQSLPD 465
QY 370 FHSKYHYQKAVVHRMQASHGEHFHVLTYLTDGTIHKVBERGEDEHSAFNIMEIQERR 429
Db 466 SEPVANLQAPQAGAVFEGSGGIRVPRANCYSES-CVDCVLRDPHCAMDESRICS 524
QY 430 AAATQMSLDARRLKLYVSSQWEVQVPLDCEYVGGGCHGLMSRDPYCGMD-QGR-CI 487
Db 525 LLSGSTKRWKMDMERNGENWCTGSPMARSPRPSPOLIKVEYLTVPSLIELACPHLSA 584
QY 488 SIYSERSVLOSINPAEHKECP-NP-K-PDK-AP-LQKVSLA-PNSRYLSCPMESR 539
Db 585 LASYHMSGRAKISEASATVYNGSL-LLLPQDGVGLY-Q--CVATENGYSYPVASYV 638
QY 540 HATYSWRH-KENVEGSCERGHQSPNCILFIENLTAQYGHFCEAGQESY-FRAGQHW 595

RESULT 13
ENTRY E42521 #type complete
TITLE A39R protein - vaccinia virus (strain Copenhagen)
ORGANISM #formal_name vaccinia virus
#note host Homo sapiens (man)
DATE 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 08-Apr-1994
ACCESSIONS E42521
REFERENCE E42521
#authors Johnson, G.P.
#submission submitted to Genbank, June 1990
#accession E42521
#status preliminary
#molecule_type DNA
#residues 1-403 #label JOH
SUMMARY #length 403 #molecular-weight 45741 #checksum 8167
Query Match 10.5%; Score 470; DB 2; Length 403;
Best Local Similarity 33.5%; Pred. No. 4.26e-74;
Matches 86; Conservative 59; Mismatches 95; Indels 17; Gaps 16;
Db 77 LVCCGNGNPNPCWK-IDSSDDPKHGRGAPYQNSKYTIISYN-ECVLSIDINISK-EG-I 132
QY 109 LACGTNAHHPSCMWLNVTVPVPLEMKGTAFFSPDENSLVLFEGDEYVSTIRKQEVNGKI 168
Db 133 KRWRRFDGPGCYDYTDADNVIPKDG-LRGAFVDRKGYD-KVYLTFDTIGSKR-I-VK 187
QY 169 PRFRIRIGE-S-ELYTSDTVNQNPQIFKATIVHDDQAVDCKIYFFREDNPDKNEAPL 226
Db 188 IPTAOKCLNDEGGPSSLSHRWSTFLKVELEC-DID-GRSY-R-Q-IHSRTIKTD-ND 241
QY 227 VSRVAQLCRDGGGSSLSVSKWTFELKAMLYCSDAATNKNFNLOVFLRDPSCGWRD 286
Db 242 TILVYFDPSPYSKALCYKSTFKOSFTSKLGYRKLQSPAPGICLPAGKAVYSTTF 301
QY 287 TRVIGVSNPNWYSAVCYVSLGIDIDKVFRTSSLGKIHSSLPNRPCKLPDQPIPTET 346

THIS PAGE BLANK (USPTO)

 Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 (TM)

Run on: Thu Jul 8 19:10:19 1999; Maspar time 18.52 Seconds
 Tabular output not generated. 914.313 Million cell updates/sec

Title: >US-09-041-236-2
 Description: (8-606) from US09041236.pep (31 of 45)
 Perfect Score: 4461
 Sequence: 1 AAASAGHLRSGPRIFAVWK.....SYFREAHQWLLPEDGIMAE 599

Scoring table: PAM 150
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 50.315; Variance 76.016; scale 0.662

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	470	10.5	403	1	VA39_VACCC PROTEIN A39.	3.04e-86
2	463	10.4	441	1	VA39_VACCV PROTEIN A39.	1.93e-84
3	163	3.7	1871	1	SEX HUMAN TRANSMEMBRANE PROTEIN	9.57e-13
4	105	2.4	227	1	US08_HCMVA HYPOTHEICAL PROTEIN H	3.93e-02
5	102	2.3	275	1	NR12_RAT T-CELL ECTO-ADP-RIBOSY	1.15e-01
6	99	2.2	275	1	NR11_RAT T-CELL ECTO-ADP-RIBOSY	3.27e-01
7	99	2.2	460	1	EF1A_TRIE ELONGATION FACTOR 1-AL	3.27e-01
8	99	2.2	591	1	VRP2_SALTJ 65 KD VIRULENCE PROTEI	3.27e-01
9	99	2.2	591	1	VRP2_SALTJ 65 KD VIRULENCE PROTEI	3.27e-01
10	99	2.2	591	1	VRP2_SALTJ 65 KD VIRULENCE PROTEI	3.27e-01
11	99	2.2	591	1	VRP2_SALTJ 65 KD VIRULENCE PROTEI	3.27e-01
12	99	2.2	593	1	VRP2_SALTJ 65 KD VIRULENCE PROTEI	3.27e-01
13	99	2.2	916	1	PM41_AJECU PLASMA MEMBRANE ATPASE	3.27e-01
14	95	2.1	1132	1	DNBI_HSV6U MAJOR DNA-BINDING PROT	4.61e-01
15	95	2.1	255	1	HA21_HUMAN HLA CLASS II HISTOCOMP	1.26e+00
16	94	2.1	329	1	RBSR_ECOLI RIBOSE OPERON REPRESSO	1.76e+00
17	93	2.1	380	1	DP9B_MICFN DNA POLYMERASE III, BE	2.44e+00
18	92	2.1	442	1	CEM1_YEAST 3-OXOACYL-(ACYL-CARRIE	3.37e+00
19	92	2.1	460	1	EF1A_NEUCR ELONGATION FACTOR 1-AL	3.37e+00
20	95	2.1	518	1	VI2_HPV5B MINOR CAPSID PROTEIN L	1.26e+00
21	95	2.1	683	1	VI2_HPV05 MINOR CAPSID PROTEIN L	1.26e+00
22	93	2.1	827	1	Y223_METUA HISTAMINE OXIDASE (EC	2.44e+00
23	94	2.1	947	1	PM42_YEAST PLASMA MEMBRANE ATPASE	1.76e+00

RESULT	ID	VA39_VACCC	STANDARD	PRT	403 AA.
AC	P21062				
DT	01-FEB-1991 (REL. 17, CREATED)				
DT	01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)				
DT	01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)				
DE	PROTEIN A39.				
GN	A39.				
OS	VACCINIA VIRUS (STRAIN COPENHAGEN).				
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXYIRIDAE; CHORPOPOXVIRINAE;				
OC	ORTHOPOXVIRUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 91021027.				
RA	GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,				
RA	PROLETTI E., "The complete DNA sequence of vaccinia virus."				
RT	VIROLOGY 179:247-266(1990).				
RL	[2]				
RP	COMPLETE GENOME.				
RA	GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,				
RA	PROLETTI E.,				
RL	VIROLOGY 179:517-563(1990).				

Query Match	Best Local Similarity	Score	DB 1: Length	DB 1: 403
Matches	86;	Conservative	59;	Mismatches 95; Indels 17; Gaps 16;
Db	77	LVCGTNGNPKCK-IGSDPKHGRGAPYQNSYTIISYN-ECVLSIDINISK-EG-I 132		
Qy	109	LACGTNRHPSCMVLGVGTPLGEMRGYAPFSDENSIVLFEDDEVYSTIRKQYNGKI 168		
Db	133	KRMRFDPGCGYDLYTDNVPKDG-LRGAFVYKDGTYD-KVYLLFDITGSKR--I-VK 187		
Qy	169	PRFRIRIGE-S-ELYTSDTWQNPQFIKATIVHODQAYDDKIYFFREDNDPKPEAPLN 226		

```

Db      168 IPIYIAOMCLNDGCGSSLSHHRWSTFLKAELECC-DID-CRSV-R-Q-IITHSTITTD-ND 241
        ::::::::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy      227 VSRVAOLCRGDGGSSSLVSXKWNFFLKMMLCSDAATKNRRLDDVFLLPDPSPGWMD 286
        ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db      242 TLIVAFEDPSYSKSLACTYSMTNTIOSFESTLEGYTKLOLPSPADICLPACKVSHHTFE 301
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy      287 TRIVGVFSNPMNYSAACYISLGDIKVFTSTLSKGTHSSLPNRPBCKCLPDQQPIPTETE 346
        ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db      302 EVIEKYNYLDDIRKLPLS 318
        ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy      347 QVADRHPVEYAQRVEDPMG 363
        ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 2
ID VA39_VACCV STANDARD: PRT: 441 AA.
AC P24764:
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DE 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DN PROTEIN A39.
GN A39R OR (SALL9R AND SALLFR).
OS VACCINIA VIRUS (STRAIN WR).
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXYRINAE;
   ORTHOPOXYVIRUS.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 91310644.
RA AMEGADDIE B.Y., AHN B.-Y., MOES B.;
RT "Identification, sequence, and expression of the gene encoding a Mr
RT 35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase.";
RL J. BIOL. CHEM. 266:13712-13718(1991).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE: 91259063.
RA SMITH G.L., CHAN Y.S., HOWARD S.T.;
RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RT the right inverted terminal repeat.";
RL J. GEN. VIROL. 72:1349-1376(1991).
RC -! CAUTION: IN REF.2 THIS SEQUENCE IS IN TWO PART (SALL9R AND
   SALLFR) DUE TO THE PRESENCE OF FRAMESHIFT COMPARED TO THAT OF
   REF.1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61187; G335798; -.
DR EMBL: D11079; G322276; ALT_SEQ.
DR EMBL: X57318; G62254; -.
DR PIR: S29921; S29921.
SQ SEQUENCE 441 AA; 50185 MM; 2C823A68 CRC32;

Query Match 10.4%; Score 463; DB 1; Length 441;
Best Local Similarity 33.9%; Pred. No. 1,93e+84;
Matches 87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;

```

Db	280	TLILYFEDSPYKSA	LCITYSANTIKQSEF	SKLEGGYTKOLPSP	PAGICLCIPACKVYPHTF	339
Qy	287	TEVYGFSNPMNVY	SAVCYSLG	IDIKVFERTSLKGYHSL	PNPRGCKLPDOOP	PTETEF 346
Db	340	EVIEKYNVLDI	IKPLS	356		
Qy	347	QVADRPEVAQ	VEPVG	363		
RESULT	3					
ID	SEX	HUMAN	STANDARD:	PRT:	1871 AA.	
AC	P51805:					
DT	01-OCT-1996	(REL. 34, CREATED)				
DT	01-OCT-1996	(REL. 34, LAST SEQUENCE UPDATE)				
DT	01-OCT-1996	(REL. 34, LAST ANNOTATION UPDATE)				
DE	TRANSMEMBRANE PROTEIN SEX PRECURSOR.					
DE	SEX.					
OS	HOMO SAPIENS (HUMAN).					
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;					
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.					
RM	(1)					
RP	SEQUENCE FROM N.A.					
RC	TISSUE-FETAL BRAIN, SKELETAL MUSCLE, AND EMBRYO;					
RX	MEDLINE: 96149362.					
RA	MAESTRINI E., TAMAGNONE L., LONGATI P., CREMONA O., GULISANO M.,					
RA	BIONE S., TAMANINI F., NEEL B.G., TONIGLO D., COMOGGIO P.M.;					
RT	"A family of transmembrane proteins with homology to the					
RT	MET-hepatocyte growth factor receptor."					
RL	PROC. NATL. ACAD. SCI. U.S.A. 93:674-678(1996).					
CC	-1- FUNCTION: PUTATIVE RECEPTOR INVOLVED IN THE DEVELOPMENT OF NEURAL					
CC	AND EPITHELIAL TISSUES.					
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).					
CC	-1- TISSUE SPECIFICITY: PREDOMINANT IN NEURAL TISSUE DURING					
CC	DEVELOPMENT (BY SIMILARITY); WIDELY EXPRESSED IN ADULT TISSUE.					
CC	-1- SIMILARITY: WITH THE EXTRACELLULAR DOMAIN OF THE MET/ROM/HGF					
CC	RECEPTORS AND THE CYTOPLASMIC DOMAIN OF SEP, NOV AND OCT.					
CC	-----					
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL, X87852; E183847; -.					
DR	MIM: 300022; -.					
KW	RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.					
FT	SIGNAL	1	19	POTENTIAL.		
FT	CHAIN	20	1871	TRANSMEMBRANE PROTEIN SEX.		
FT	DOMAIN	20	1220	EXTRACELLULAR (POTENTIAL).		
FT	DOMAIN	1221	1241	POTENTIAL.		
FT	DOMAIN	1242	1871	CYTOPLASMIC (POTENTIAL).		
FT	CARBOHYD	59	59	POTENTIAL.		
FT	CARBOHYD	548	548	POTENTIAL.		
FT	CARBOHYD	637	637	POTENTIAL.		
FT	CARBOHYD	738	738	POTENTIAL.		
FT	CARBOHYD	746	746	POTENTIAL.		
FT	CARBOHYD	1009	1009	POTENTIAL.		
FT	CARBOHYD	1036	1036	POTENTIAL.		
FT	CARBOHYD	1073	1073	POTENTIAL.		
FT	CARBOHYD	1115	1115	POTENTIAL.		
FT	CARBOHYD	1162	1162	POTENTIAL.		
SO	SEQUENCE	1871 AA;	207661 MW;	9A11046A CRC32;		
Query Match	3.7%;	Score 163;	DB 1;	Length 1871;		
Best Local Similarity	29.5%;	Pred. No. 9,576-13;				
Matches	26;	Conservative	29;	Mismatches 29;	Indels 4;	Gaps 3;
Db	427	VVFICTRSGSLKRV	VDGFQDAHL-YETVPYVD--GSPILRDLLFS	PDHRIHYLISEKOV	483	
Qy	394	VLYLTDTGTHK	WVPEQESHFENFMELQEPFRAA	IQTMSLDARRLKLYSSQWEV	453	

[illegible]

```

OS RATTUS NORVEGICUS (RAT).
OC EURAROTTA METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA.
RN ROBERTIA; SCUROGNATHI; MORIDA; MORINAE; RATTUS.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 9013897.
RA KOCH F., HAAG F., KASHAN A., THIELE H.-G.:
RT "Primary structure of rat Rt6.2, a nonglycosylated
RL phosphatidylinositol-linked surface marker of postthymic T cells."
RM PROC. NATL. ACAD. SCI. U.S.A. 87:964-967(1990).
RN [2]
RP SEQUENCE OF 1-201 FROM N.A.
RC STRAIN-DA; TISSUE=SPLEEN;
RX MEDLINE: 96355006.
RA HAAG F., KUHNENBUMER G., KOCH-NOLTE F., WINGENDER E., THIELE H.-G.:
RT "Structure of the gene encoding the rat T cell ecto-ADP-
RL ribosyltransferase Rt6."
RM J. IMMUNOL. 157:2022-2030(1996).
RN [3]
RP PARTIAL SEQUENCE.
RX MEDLINE: 90215814.
RA KASHAN A., BOCK F., HAAG F., KOCH F., THIELE H.-G.:
RT "A single-step purification procedure and partial amino acid sequence
RL analysis of picomole amounts of the rat T cell alloantigen Rt6.2."
RN IMMUNOL. LETT. 23:133-138(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE: 94193612.
RA TAKADA T., IIDA K., MOSS J.:
RT "Expression of NAD glycohydrolase activity by rat mammary
RL adenocarcinoma cells transformed with rat T cell alloantigen Rt6.2.";
RJ BIOC. CHEM. 269:9420-9423(1994).
CC - FUNCTION: HAS BOTH NAD+ GLYCOHYDROLASE AND ADP-RIBOSYLTRANSFERASE
CC ACTIVITY (TO A LESSER EXTENT)
CC - CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
CC N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC - CATALYTIC ACTIVITY: NAD(+) + H(2O) = NICOTINAMIDE + ADP-RIBOSE.
CC - SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC - TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC - PFM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC - SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M85193; G206806; -
DR EMBL: X99123; E257753; -
DR EMBL: X99122; E257751; -
DR PIR: A34866; A34866.
DR PROSITE: PS01291; ART; 1.
DR PFAM: PF01129; ART; 1.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NMD; SIGNAL;
KW T-CELL DIFFERENTIATION; GPI-ANCHOR.
FT SIGNAL 1 20
FT CHAIN 21 246 T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 2.
FT PROPEP 247 275 HYDROPHOBIC, REMOVED DURING MATURATION
FT FT BY SIMILARITY.
FT LIPID 246 246 GPI-ANCHOR (BY SIMILARITY).
FT ACT_SITE 209 209 BY SIMILARITY.
SO SEQUENCE 275 AA; 31438 MW; DEBA84E CRC32;

```

OY 131 LGE-MRGAFSPDENSILVLEFGDEVYSTIRKQENCK-IPFRIRIRGESE-LYTSOTVM 187
Db 252 ES 253
OY 188 ON 189

RESULT 6
ID NRT1-RAT STANDARD: PRT: 275 AA.
AC P17982;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 PRECURSOR (EC 2.4.2.31) (T-CELL
NAD(P)(+)-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL MONO(ADP-
RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN
RT6.1).
GN ART2A OR RT6-A.
OS RATTUS NORVEGICUS (RAT)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIURIONATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEWIS A;
RX MEDLINE: 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell
RT allantoigen RT6.1".
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
RN [2]
RP MUTAGENESIS OF GLN-207.
RX MEDLINE: 96275529.
RA MACHAMA T., HOSHINO S.-I., KATADA T.;
RT "Increase in ADP-ribosyltransferase activity of rat T lymphocyte
RT allantoigen RT6.1 by a single amino acid mutation.";
RL FEBS LETT. 388:189-191(1996).
CC -1- FUNCTION: HAS NAD+ GLYCOHYDROLASE ACTIVITY AND EXTREMELY LOW ADP-
CC RIBOSYLTRANSFERASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE -> NICOTINAMIDE +
CC N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -1- CATALYTIC ACTIVITY: NAD(+) + H(2)O -> NICOTINAMIDE + ADP-RIBOSE.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -1- PM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLY).
CC -1- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52082; G57168; -;
DR EMBL: M31138; G206804; -;
DR PIR: S08464; S08464;
DR PROSITE: PS01291; ART: 1.
DR PFAM: PF01129; ART: 1.
KW TRANSFERASE: GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
KW T-CELL DIFFERENTIATION; GPI-ANCHOR.
FT SIGNAL 1 20
FT CHARIN 21 246 T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
FT PROPEP 247 275 HYDROPHOBIC, REMOVED DURING MATURATION
FT (BY SIMILARITY).
FT LIPID 246 246 GPI-ANCHOR (BY SIMILARITY).
FT ACT_SITE 209 209 POTENTIAL.
FT CARBOHYD 58 58 BY SIMILARITY.
FT MUTAGEN 207 207 Q->E: INCREASED ADP-RIBOSYLTRANSFERASE
FT ACTIVITY.
SQ SEQUENCE 275 AA: 31388 MW: 52318464 CRC32:

Query Match 2.2%; Score 99; DB.1; Length 275;
Best Local Similarity 30.6%; Pred. No. 3.27e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;

Db 194 LGVYKEFS-FYPOOE-VLIPGEYVOKVKTQGVNEFLDSPPKRSNVCYSSAGTR 251
OY 131 LGE-MRGAFSPDENSILVLEFGDEVYSTIRKQENCK-IPFRIRIRGESE-LYTSOTVM 187
Db 252 ES 253
OY 188 ON 189

RESULT 7
ID EF1A-TREIR STANDARD: PRT: 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEPI.
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EURASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-QM9414 / RUTC-30;
RX MEDLINE: 94124021.
RA NAKARI T., ALATALO E., PENTTILA M.;
RT "Isolation of Trichoderma reesei genes highly expressed on glucose-
RT containing media: characterization of the tef1 gene encoding
RT translation elongation factor 1 alpha.";
RL GENE 136:313-318(1993).
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TV/EF-1A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z23012; G312887; -;
DR PIR: S35772; S35772;
DR PROSITE: PS00301; EFATOR-GTP: 1.
DR PFAM: PF00009; GTP_EFTU; 1.
DR HSSP: P07157; 1AIP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA: 49830 MW: B9ABAB82 CRC32:

Query Match 2.2%; Score 99; DB.1; Length 460;
Best Local Similarity 33.3%; Pred. No. 3.27e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;

Db 210 YKGEKTRKAGTG-KTLEAIDISIEPPKR-PTDKPLRLPLQDV 252
OY 478 YCGMDGRCISYSSERSVLOSINPAEPHKCPKPKAPLQAV 522

RESULT 8
ID VRP2-SALTY STANDARD: PRT: 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN MKAA.
OS SALMONELLA TYPHIMORIUM.
OC PLASMID 96 KB VIRULENCE PEX102.
CC BACTERIA: PROTEOBACTERIA: GAMMA SUBDIVISION: ENTEROBACTERIACEAE;
CC SALMONELLA.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-TML R66;
RX MEDLINE: 90136009.
RA TAIIRA S., RHEN M.;
RT "Identification and genetic analysis of mkaa -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth."
RL MICROB. PATHOG. 7:165-173(1989).
RN (2)
RP SEQUENCE OF 1-10.
RX MEDLINE: 91244158.
RA TAIIRA S., BAUDANN M., RIKKONEN P., SUKUPOLVI S., RHEN M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium."
RL FEMS MICROBIOL. LETT. 61:319-323(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 215042; G47783; -
DR PIR: A54540; A54540.
KW PLASMID: VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65440 MW; B69F25C3 CRC32;

Query Match 2.2%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.27e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

DB 162 HDSNGILHLGKTAARLSDPOAASHTAOW-LVEESVTPAGE-HIYSYLAENGVDN 219
QY 92 RDCENVITLLERSEGLACGTNARHPSCMNLVNGTVPLGEMRGYAPSPDENSIVLFE 151
DB 220 GNEAGRDRSAMRYLSKV-OYGNATPAADLY 248
QY 152 GDEVSTIRKOEYNGKIPFRIRIGESELY 181

RESULT 9
ID VRP2_SALCH STANDARD: PRT: 591 AA.
AC P17450:
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN (PROTEIN M2, IN MBA REGION).
OS SALMONELLA CHOLERAE-SUIS (SALMONELLA ENTERICA).
OC PLASMID PKDSC50
CC BACTERIA: PROTEOBACTERIA: GAMMA SUBDIVISION: ENTEROBACTERIACEAE;
CC SALMONELLA.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-RF-1;
RX MEDLINE: 90245675.
RA MATSUI H.;
RT "Nucleotide sequences of genes encoding 32 kDa and 70 kDa
RT polypeptides in mba region of the virulence plasmid, pkDsc50, of

RT Salmonella choleraesuis.";
RL NUCLEIC ACIDS RES. 18:2181-2181(1990).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52035; G46898; -
DR PIR: S09498; S09498.
KW PLASMID: VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65341 MW; BE4A439A CRC32;

Query Match 2.2%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.27e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

DB 162 HDSNGILHLGKTAARLSDPOAASHTAOW-LVEESVTPAGE-HIYSYLAENGVDN 219
QY 92 RDCENVITLLERSEGLACGTNARHPSCMNLVNGTVPLGEMRGYAPSPDENSIVLFE 151
DB 220 GNEAGRDRSAMRYLSKV-OYGNATPAADLY 248
QY 152 GDEVSTIRKOEYNGKIPFRIRIGESELY 181

RESULT 10
ID VRP2_SALEN STANDARD: PRT: 591 AA.
AC P55220:
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN SPVB.
OS SALMONELLA ENTERITIDIS.
OC PLASMID PNL2001.
CC BACTERIA: PROTEOBACTERIA: GAMMA SUBDIVISION: ENTEROBACTERIACEAE;
CC SALMONELLA.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-AL1190;
RX MEDLINE: 94362897.
RA SUZUKI S., KOMASE K., MATSUI H., ABE A., KAWAHARA K., TAMURA Y.,
RA KITJMA M., DANBARA H., NAKAMURA M., SATO S.;
RT "Virulence region of plasmid PNL2001 of Salmonella enteritidis.";
RT MICROBIOLOGY 140:1307-1318(1994).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D14490; G517164; -
DR PIR: S09498; S09498.
KW PLASMID: VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65350 MW; 9C5B2EE3 CRC32;

Query Match 2.2%; Score 99; DB 1; Length 591;

[illegible]

```

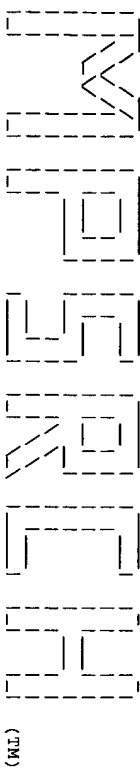
OC ONYGENALES: ONYGENACEAE, ADELLOWICES.
RN [1]
RN SEQUENCE FROM N.A.
RA MEDLINE: 94124018.
RX SCHAEFER M.P., DEAN G.E.:
RT "Cloning and sequence analysis of an H(+) -ATPase-encoding gene from
RL the human dimorphic pathogen Histoplasma capsulatum.",
RL GENE 136:295-300(1993).
CC -I- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING
CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKALINIZATION MAY MEDIATE
CC GROWTH RESPONSES.
CC -I- CATALYTIC ACTIVITY: ATP + H(2)O -> ADP + ORTHOPHOSPHATE.
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L07305; G409249; -.
CC DR PROSITE: PS00154; ATPASE_E1-E2; 1.
CC PFAM: PF00122; E1-E2_ATPase; 1.
CC KM HYDROLASE: HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
CC ATP-BINDING.
CC FT DOMAIN 1 111 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 112 134 1 (POTENTIAL).
CC FT DOMAIN 135 136 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 137 155 2 (POTENTIAL).
CC FT DOMAIN 156 279 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 280 305 3 (POTENTIAL).
CC FT DOMAIN 306 320 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 321 350 4 (POTENTIAL).
CC FT DOMAIN 351 711 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 712 734 5 (POTENTIAL).
CC FT DOMAIN 735 754 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 755 776 6 (POTENTIAL).
CC FT DOMAIN 777 817 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 818 843 7 (POTENTIAL).
CC FT DOMAIN 844 847 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 848 874 8 (POTENTIAL).
CC FT DOMAIN 875 916 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 916 916 ASP/GLU-RICH (ACIDIC).
CC FT MOD_RES 374 374 PHOSPHORYLATION (BY SIMILARITY).
CC FT BINDING 470 470 ATP (BY SIMILARITY).
CC SQ SEQUENCE 916 AA; 98884 MM; 4446844 CRC32;

Query Match 2.2%; Score 99; DB 1; Length 916;
Best Local Similarity 37.2%; Pred. No. 3; 276-01;
Matches 16; Conservative 8; Mismatches 16; Indels 3; Gaps 3;

DB 762 LAVGTWIT-LITMLVSGNGGIYVNGRTHPYLFLIEISLIEW 803
QY 244 LSVSKWNTFLKAMLVCS-D-AATNKNFNLDV-FLPDPSPGOW 284

RESULT 13
ID DNBI_HSV6U STANDARD: PRT: 1132 AA.
AC P52338.
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MAJOR DNA-BINDING PROTEIN (MBDP).
GN UAI.
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UCANDA-1102).
OC VIRUSES: DSDNA VIRUSES, NO RNA STRAIN: HERPESVIRINAE;
OC BETAHERPESVIRINAE; ROSEOLOVIRUS.

```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_dp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 19:10:59 1999; MasPar time 37.55 Seconds
870.595 Million cell updates/sec
Tabular output not generated.

Title: >US-09-041-236-2
Description: (8-606) from US09041236.pep (31 of 45)
Perfect Score: 4461
Sequence: 1 AASAGHLRSGPRIFAVWK.....SYFREAQHKQLLPEDGIMAE 599

Scoring table: PAM 150
Gap 11
Searched: 179066 seqs, 54579741 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptribml9
1:sp-Archea 2:sp-Bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 48.787; Variance 75.124; scale 0.649
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	4461	100.0	666	4	SEMAPHORIN L.	0.00e+00
2	2354	52.8	393	11	SEMAPHORIN L. (FRAGMENT	0.00e+00
3	1862	41.7	653	14	SIMILAR TO GENBANK ACC	0.00e+00
4	737	16.5	748	11	SEMAPHORIN A PRECURSOR	1.62e-154
5	725	16.3	751	11	SEMAPHORIN E PRECURSOR	2.37e-151
6	716	16.1	749	4	SEMAPHORIN V.	5.59e-149
7	710	15.9	751	4	SEMAPHORIN E.	2.13e-147
8	705	15.8	751	13	COLLAPSED 3.	4.40e-146
9	688	15.4	753	4	SEMAPHORIN III FAMILY	1.30e-141
10	682	15.3	754	11	SEMAPHORIN IV ISOFORM	4.89e-140
11	681	15.3	772	13	COLLAPSED.	8.96e-140
12	682	15.3	785	11	SEMAPHORIN IV ISOFORM	8.96e-140
13	681	15.3	785	4	SEMAPHORIN IV	3.00e-139
14	679	15.2	785	4	SEMAPHORIN D PRECURSOR	1.26e-136
15	679	15.0	772	11	SEMAPHORIN III/COLLAPD	1.57e-134
16	668	15.0	772	11	SEMAPHORIN C (SEM C) (2.97e-122
17	668	15.0	772	11	SEMAPHORIN H.	
18	661	14.8	771	4		
19	618	13.9	782	11		
20	614	13.8	775	11		

21	599	13.4	861	11	SEMAPHORIN J (SEMAPHOR	2.37e-118
22	594	13.3	761	13	COLLAPSED-2.	4.71e-117
23	592	13.3	775	4	KIAA0331.	1.56e-116
24	572	12.8	785	13	COLLAPSED 5.	2.38e-111
25	568	12.7	294	13	COLLAPSED-3 (FRAGMENT)	2.58e-110
26	544	12.2	862	4	SEMAPHORIN I	4.04e-104
27	522	11.7	834	11	SEMAPHORIN I (M-SEMA F	1.83e-98
28	511	11.5	299	13	COLLAPSED-5 (FRAGMENT)	1.21e-95
29	480	10.8	760	11	SEMAPHORIN B PRECURSOR	9.88e-88
30	473	10.6	795	13	COLLAPSED-4 (FRAGMENT)	5.94e-86
31	461	10.3	730	5	FASCICLIN IV.	6.57e-83
32	448	10.0	1074	4	SEMAPHORIN F HOMOLOG.	1.27e-79
33	441	9.9	1077	11	SEMAPHORIN F PRECURSOR	7.40e-78
34	416	9.3	1093	11	SEMAPHORIN G PRECURSOR	1.40e-71
35	392	8.8	888	11	SEMAPHORIN VIA.	1.34e-65
36	390	8.7	494	4	SEMAPHORIN F (FRAGMENT	4.19e-65
37	378	8.5	712	5	SEMAPHORIN-I PRECURSOR	3.91e-62
38	367	8.2	562	5	CESEMA.	2.01e-59
39	367	8.2	771	5	SEMAPHORIN-I.	2.01e-59
40	362	8.1	706	5	SEMAPHORIN-II.	3.40e-58
41	341	7.6	284	11	SEMAPHORIN IV HOMOLOG	4.59e-53
42	316	7.1	887	11	SEMAPHORIN 2 (SEMAPHOR	5.05e-47
43	316	7.1	887	11	SEMAPHORIN 2 (SEMAPHOR	5.05e-47
44	283	6.3	770	5	SIMILAR TO SEMAPHORIN-	3.55e-39
45	204	4.6	1963	4	KIAA0463 PROTEIN (FRAG	3.19e-21

ALIGNMENTS

RESULT 1
ID 075326 PRELIMINARY; PRT; 666 AA.
AC 075326;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN L.
GN SEMAL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
viruses";
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030698; G3523115; -;
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match	100.0%	Score 4461;	DB 4;	Length 666;
Best Local Similarity	100.0%	Pred. No. 0.00e+00;		
Matches	599;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Db	40	AASAGHLRSGPRIFAVWKGVGDVDFGOTEPHTVLFHEPGSSSWVGGRKYLFD	99	
Qy	8	AASAGHLRSGPRIFAVWKGVGDVDFGOTEPHTVLFHEPGSSSWVGGRKYLFD	67	
Db	100	FPEKKNASVTVNIGSTKSGCLDKDCENYITLLERSEGLACGTNARHPSGMLVNGT	159	
Qy	68	FPEKKNASVTVNIGSTKSGCLDKDCENYITLLERSEGLACGTNARHPSGMLVNGT	127	
Db	160	VVPLGEMRGYAPSPDNLVLEGGDVYSTIRKQENKGIPEFRIRGRGSELYTSPTVM	219	
Qy	128	VVPLGEMRGYAPSPDNLVLEGGDVYSTIRKQENKGIPEFRIRGRGSELYTSPTVM	187	
Db	220	ONPQFIKATVHODQAVDDKIYFFREDNDPKNDEAPLANSPVAOLCRGOGGESSLSVS	279	
Qy	188	ONPQFIKATVHODQAVDDKIYFFREDNDPKNDEAPLANSPVAOLCRGOGGESSLSVS	247	
Db	280	KWNTFLKAMLVCSDAATNKNFNRLQDVFLLPDPSGQWRDTRVYGVSNPNYSAVCYSL	339	
Qy	248	KWNTFLKAMLVCSDAATNKNFNRLQDVFLLPDPSGQWRDTRVYGVSNPNYSAVCYSL	307	

Db	340	GGDKVFRSSSLKGYHSSLPNRPQGCPLDDOIPETETOVADRHAEVQAVRYPMPKLT	399
Oy	308	GIDIKVFRSSSLKGYHSSLPNRPQGCPLDDOIPETETOVADRHAEVQAVRYPMPKLT	367
Db	400	PLEHSKYHYQKVAVHRMQASHGETFFVLVLYLTDRGTHHKVPEGEDEHSFAFNIMEIOPE	459
Oy	368	PLEHSKYHYQKVAVHRMQASHGETFFVLVLYLTDRGTHHKVPEGEDEHSFAFNIMEIOPE	427
Db	460	RRAAAIQTMSLDAERKKLVYSSQMEVSOYPLDCEVYGGGCHCLMSRDPYCGMDGRCI	519
Oy	428	RRAAAIQTMSLDAERKKLVYSSQMEVSOYPLDCEVYGGGCHCLMSRDPYCGMDGRCI	487
Db	530	SIYSSERSVYLOSINPAEPHKECPNPKPDPAPLQKYSIAENSRYYLSCPMESRATYTSWRH	579
Oy	488	SIYSSERSVYLOSINPAEPHKECPNPKPDPAPLQKYSIAENSRYYLSCPMESRATYTSWRH	547
Db	580	KENVSOCSGPGHOSPPICLFIENLTAAOQGYHVFCEAOBESYFPEAAHWOLLPEDGIMAE	638
Oy	548	KENVSOCSGPGHOSPPICLFIENLTAAOQGYHVFCEAOBESYFPEAAHWOLLPEDGIMAE	606

RESULT	2		
ID			
AC	088371	PRELIMINARY:	PRT: 393 AA.
AD	088371:		
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)	
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)	
DE	SEMAPHORIN L (FRAGMENT).		
GN	SEMAP.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUMAROTIA, MEALAZOA, CHORDATA; VERTEBRATA, MAMMALIA, EUTHERIA; RODENTIA		
OC	SCIROGNATHI; MORIDAE; MURINAE; MOS.		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 98389619.		
RA	"LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;		
RT	"NEW eukaryotic semaphorins with close homology to semaphorins of DNA		
RL	viruses" 51:340-350(1998).		
GENB	EMBL, AF030699; G3523117; -.		
DR	NON_TER 393		
QO	SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;		

Query Match	52.8%	Score 2354	DB 11	Length 393
Best Local Similarity	91.0%	Pred. No. 0.00e+00		
Matches 355; Conservative	14;	Mismatches 14;	Indels 4;	Gaps 2

Db	40	AAASAGCSBSRGSPRTSAVMKG---QDAHYDSQEPPEHYLFLHEPSSSEVWVGGRKRYHEN	96
Qy	8	AAASAGGHLRSGPRIFFAWMKHGVODRFGQTPPHVLYLHEPSSSSVWVGGRKYYLFD	67
Db	97	FPECGMAVSRVTVNIGSTGSCODKODCGNTYTLLEBRNGNLGCTARKPSCGNLYNDS	156
Qy	68	FPECGMAVSRVTVNIGSTGSLCDKRDCENTYTLLEBRSEGLACGTARHPSCNLYNGT	127
Db	157	VVMSSGEMKGAPESPDSNSLYLFECDGEVYSTIRKOEYNGKIPFRRIIGSELYTSDTV	216
Qy	128	VV-PLGEMRGAPSPSPDSNSLYLFECDGEVYSTIRKOEYNGKIPFRRIIGSELYTSDTV	166
Db	217	MONPOFIKATIVHODQAYDCKIYFFREDNDKKNPEAPLNVSRVAOLCRDQOGESSLSV	276
Qy	187	MONPOFIKATIVHODQAYDCKIYFFREDNDKKNPEAPLNVSRVAOLCRDQOGESSLSV	246
Db	277	SKMWTFLKAMLYCSDAATNNRNFNLQDVEFLLPDSGQMRTRYGVFSNPMNTSAVCVYS	336
Qy	247	SKMWTFLKAMLYCSDAATNNRNFNLQDVEFLLPDSGQMRTRYGVFSNPMNTSAVCVYS	306
Db	337	LGGIDLRVFRTSLSLKGYNHGLSNRPFGMCLPKKKQIPIPETQYVADSHFEVYAQRVEPMG	393
Qy	307	LGGIDLRVFRTSLSLKGYNHGLSNRPFGMCLPKKKQIPIPETQYVADSHFEVYAQRVEPMG	363

ID	RESULT	3	PRELIMINARY;	PRT;	653	AA.
AC		064906.				
AD		01-NOV-1996	(TREMBLREL. 01,	CREATED)		
DT		01-NOV-1996	(TREMBLREL. 01,	LAST SEQUENCE UPDATE)		
DT		01-NOV-1998	(TREMBLREL. 08,	LAST ANNOTATION UPDATE)		
DE			SIMILAR TO GENBANK ACCESSION NUMBER	L26081.		
OS			ALCELAAPHINE HERPESVIRUS 1.			
OC			VIRUSES: DSDNA VIRUSES.	NO RNA STAGE; HERESVIRIDAE:		
OC			GAMMAHERPESVIRINAE.			
RN			[1]			
RP			SEQUENCE FROM N.A.			
RC			STRAIN-CS00;			
RX			MEDLINE: 97201573.			
RA			ENSER A., FLECKENSTEIN B.;			
RT			"Alcelaphine herpesvirus type 1 has a semaphorin-like gene."			
RU			J GEN. VIROL. 76:1063-1067(1995).			

RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RX MEDLINE: 97404659.
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alphacellulose herpessavirus 1 genome."; J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-C500;
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U18243; GI000117; -.
DR EMBL: AF005370; G2337970; -.
SO SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match	41.78;	Score 1862;	DB 14;	Length 653;
Best Local Similarity	46.38;	Pred. No. 0.00e+00;		
Matches	211;	Conservative 102;	Mismatches 200;	Indels 12;
			Gaps 12;	

[illegible]

DB 596 YFPTKGTETKRGVHTKNDCLLLANSTANGTHVCMKEDS 640
QY 543 YSMRKENVEQSCPEGHSPNCILFIENLTAQOYGHVCEAEGS 587
RESULT 4
ID 062177 PRELIMINARY; PRT; 748 AA.
AC 062177;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN A PRECURSOR (SEM A).
GN SEMA OR SEMA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SC CINOGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO. LOW LEVELS FOUND BETWEEN DAYS 10-12.
CC EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL BIRTH.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85990; G854324; -.
DR MGD; MGI:107561; SEMA.
DR PFAM; PF00047; 19; 1.
KW SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KM DEVELOPMENTAL PROTEIN.
FT CHAIN 1 26
FT SIGNAL 27 748
FT DOMAIN 586 649
FT SEQUENCE 748 AA: 82894 MW; A7E53ABD CRC32;
Query Match 16.5%; Score 737; DB 11; Length 748;
Best Local Similarity 33.7%; Pred. No. 1.62e-154;
Matches 168; Conservative 112; Mismatches 166; Indels 52; Gaps 36;
DB 112 ECANFVRLAHAYNHTHLLACRTGAFHPTCALMKRATAGGTHASTGPEKLEDGSKTYPD 171
QY 93 DCENYITLLRERSEG-LLAGGTNARHPSG---WNLVNGT-V-V-P--LGEHROYAFESP 142
DB 172 RHRPPSVLVEELYSGV-TADLMGRDFTFRSGQNPSTLTPHDSWMLPEPKVYKFWI 230
QY 143 DENSLVLFEDSEYSTIRKOEYNGK-IPRRIRGESELT-S-DIV-MONPOFIKT-I 197
DB 231 PESENDDDKIYFFFEESAEAAAPAMGMSVSRGQICRNDLGQRL-VNKKTTPEKAR 289
QY 198 VHDQADAKIYFFFRDNDKNPE-APLNVSRVQACRDOGSESLSVSKWTFIKAM 256
DB 290 LVCSVPEGEDTHFDQADVLLS-SR-DROTPLLAVFSTSGVFGSACVYISMDVR 347
QY 257 LVCS--DAANKNFNRLQDVFLLPDPGQWDRTRVYGVFSNPWN-Y--SAVCVSLGDID 311
DB 348 RATLGLPRHKEGPTHQWVSQGRVPRPGMCPSTKGTSSSTKDFDDVYIQGRNPLM 407
QY 312 KVERTS-SLK-G-----HSSLPNRPCKCLPDD-Q-PIPT-TETF--QVAD-R-RPEV 355
DB 408 YNPLPMGG-R-PLFLOVAGAGYFTQIAADRVAAADG-H-DVLEIGDVGTVLKIVSPK 464
QY 356 AQVRPEPGLKPLF-H--SKYHQAAYAHVMAQASHGETHVLTLTDRGTIRKHYVE-P- 410
DB 465 GRBNSEGLLLELQVEDSAITSMQISSRKQOLYVASRAVAQIALHRCIALGRACAE 524

QY 411 GEDEHFAFINIMEIOFRRRAAIQMSLDAERKLYVSSQWVEVQPLDCEVYGGCHG 470
DB 525 CCLARDPYCAMDSACTRQPTAKRRFRRODIRNGDPTLCSGDSHVLLEKYLGVES 584
QY 471 CLMSRDPYCGMDGRCISISYSRSVL--QSINPAEPHKECPNPKDKAPLOKVSAPNS 528
DB 585 GSAFLCEPERSLOAHYQW 602
QY 529 -RYLSCPMESRHATYSW 545
RESULT 5
ID 062181 PRELIMINARY; PRT; 751 AA.
AC 062181;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E PRECURSOR (SEM E).
GN SEMAE OR SEME.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SC CINOGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO. MAXIMUM EXPRESSION BETWEEN DAYS 10-12 WITH MODERATE LEVELS FROM DAY 13 UNTIL BIRTH.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85994; G854332; -.
DR MGD; MGI:107557; SEMAE.
DR PFAM; PF00047; 19; 1.
KW SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KM DEVELOPMENTAL PROTEIN.
FT CHAIN 1 20
FT SIGNAL 21 751
FT DOMAIN 587 649
FT SEQUENCE 751 AA: 85259 MW; B2BDCRE CRC32;
Query Match 16.3%; Score 725; DB 11; Length 751;
Best Local Similarity 33.0%; Pred. No. 2.37e-151;
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;
DB 167 SENPNVNTVSMNEELFSGMYT-DEMGIDDAFBSLRTRMQRLTROHNSKULSEPMFVD 225
QY 139 PESPENSLVLFEGDEYSTIRKOEYNGK-IPRRIRGESELYTS--DIV-MONPOFIK 194
DB 226 AHVIPGTDNDKAYFFPERLTIDNRSKTOJHSIARICPDNTGGORSL-VNKKTTPL 284
QY 195 ATIVHOD-QAYDKIYFFREDNPDKNPEAPLNVSRVQACRDOGSESLSVSKWTFIL 253
DB 285 KARLVCSVTDDEDEPETHFELEDFLL-ETDNP-RTLLVYGITTSYFVKGSAVCVYHL 342
QY 254 KAMLVCS--DA-ATNKNFNRLQDVFLLPDPGQWDRTRVYGVFSNPWN-Y--SAVCVSL 307
DB 343 SDIQYVFNCEFAHKEGPNHOLISYQGRIPYPRGTCPGCAFYTNMRTTDFDPDVYTFR 402
QY 308 GDIKVERTS-SLK-G-----YHSSLPNRPCKCLPDD-Q-PIPT-TETF--QVADRP 353
DB 403 NHPILMNSISPIHRPLIYAGTDYKTKAANDRVAAADG-RVHYLFLGDRCTGVQYV 461
QY 354 EVAQVRPEPGL-KTPLF-H--SKYHQAAYAHVMAQASHGETHVLTLTDRGTIRKHYVE 409
DB 462 LPTNSASGELLLELVEFVKHNPITTMETSSKKQOLYVSNQGVQVSLHRCIHYGTAC 521

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA YAMADA T., GOTOH M., HIROHASHI S.;
RT Identification of semaphorin E as a non-MDR drug resistance gene of
human cancers.";
RL PROC. NATL. ACAD. SCI. U.S.A. 94:14716-14718(1997).
DR EMBL: AB000220; D103360; -.
SQ SEQUENCE 751 AA; 85207 MW; F89FB934 CRC32;

Query Match 15.98; Score 710; DB 4; Length 751;
Best Local Similarity 33.68; Pred. No. 2,13e-147;
Matches 151; Conservative 104; Mismatches 151; Indels 43; Gaps 30;

Db 167 SFNPNNVTVSMINELFSGMYI-DEMGTDAAIFRSLTKRNAVRTDOHNSKMLSEPFVD 225
QY 139 PESPDSNLVLFEGDEYVSTIRKOEYNGK-IPRFRRIRGSELYTS--DIV-MQNQPFIK 194
Db 226 AHVIPDGTDPNDKAYIFFEFEKLTLDNNSRSTKQIHSIMARICPNDTGSLRSL-VNKWTFEL 284
QY 195 ATIVHOD-QAYDDKIIYFFREDNDKNPAPLNVSRVAQLCRGDGOESSLSYSKNWTFEL 253
Db 285 KALVCSVMDDEGTENYFDELEDFVLL-ETDNP-RTTLVYGITFTSSIFKSGAVCYHL 342
QY 254 KALVCS--DA-ATNKNFNLODFLLPDPSCGWRDTRVGVSNPNV-Y--SAVCYSL 307
Db 343 SDIQTFENGPFARKEGPNHOLISYOGRIPIYPRTGCPGAFTPNMRTKEFPDDVTFEIR 402
QY 308 GDIDKVFRTS-SLK-G-----YHSSLPRPRGKCLPDQ-QP-IPLET-F--QVADH-R 351
Db 403 NHPLMTNPIPIKRLPIVRI-GTDYKTKIAYDRVNAADG-RYHFLGTDRGTQKVY 460
QY 352 -HBEVAQVRPEM-G-PLKTPLEFHSKYHQVAVHMQASHGEFHVLYLTDRGTIKHYV 408
Db 461 VLPTNNSVSEELLELEVEFGKKNAPITTKISSKQOLYVSSNEGVSQVLRHCITGTA 520
QY 409 E-GEQESHFAFNIMEIQPFRRAAIIOTMSLDERRKLYVSSQWESQVPLDCEVYGGG 467
Db 521 CADCCCLARDPYCAMDGNCSRFYPTGKRSRRODVHRGNPLTCRGFNLAARYNAETVQ 580
QY 468 CHCLMSRDYPCGMDGRCISITISSE--RSVLOSINAPAEHKECP--NPKDKAPLQKVS 523
Db 581 YGVKNNTFLECAPKSPQASIKWLLQDN 609
QY 524 LAF-NSRYLSCPMESRHATYSNR-HKEN 550

RESULT 9
ID 042236 PRELIMINARY: PRT; 751 AA.
AC 042236;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COLAPSPIN 3.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGATIAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA FEINER L., KOPPEL A.M., KOBAYASHI H., RAPER J.A.;
RL NEURON 19:0-0(1997).
DR EMBL: AF022946; G2522204; -.
SQ SEQUENCE 751 AA; 85432 MW; 0B9BA2F7 CRC32;
Query Match 15.88; Score 705; DB 13; Length 751;
Best Local Similarity 33.58; Pred. No. 4,40e-146;
Matches 150; Conservative 107; Mismatches 150; Indels 41; Gaps 28;

Db 167 SFNPNNVTVSMINELFSGMYI-DEMGTDAAIFRSLTKRNAVRTDOHNSKMLSEPFVD 225
QY 139 PESPDSNLVLFEGDEYVSTIRKOEYNGK-IPRFRRIRGSELYTS--DIV-MQNQPFIK 194
Db 226 AHVIPDGTDPNDKAYIFFEFEKLTLDNNSRSTKQIHSIMARICPNDTGSLRSL-VNKWTFEL 284
QY 195 ATIVHOD-QAYDDKIIYFFREDNDKNPAPLNVSRVAQLCRGDGOESSLSYSKNWTFEL 253
Db 285 KALVCSVMDDEGTENYFDELEDFVLL-ETDNP-RTTLVYGITFTSSIFKSGAVCYHL 342
QY 254 KALVCS--DA-ATNKNFNLODFLLPDPSCGWRDTRVGVSNPNV-Y--SAVCYSL 307
Db 343 SDIQTFENGPFARKEGPNHOLIPYOGRIPIYPRTGCPGAFTPNMRTKEFPDDVTFEIR 402
QY 308 GDIDKVFRTS-SLK-G-----YHSSLPRPRGKCLPDQ-QP-IPLET-F--QVADH-R 353
Db 403 NHPLMTNPIPIKRLPIVRI-GTDYKTKIAYDRVNAADGRC-HVFLGTDRGTQKVY 461
QY 354 EVAQVRPEMGP-PL-KTPLF-H--SKHYQKVAHMQASHGEFHVLYLTDRGTIKHYVE 409
Db 462 LPTNFSASGELLELEVEFGSNPITTKISSKQOLYVSSNEGVSQVLPNHRIRYGTAC 521
QY 410 -PGEQESHFAFNIMEIQPFRRAAIIOTMSLDERRKLYVSSQWESQVPLDCEVYGGG 468
Db 522 ADCCLARDPYCAMDGNCSRFYPTGKRSRRODVHRGNPLTCRGFNLAARYNAETVQY 581
QY 469 HCLMSRDYPCGMDGRCISITISSE--RSVLOSINAPAEHKECP--NPKDKAPLQKVS 524
Db 582 YGVKNNTFLECAPKSPQASIKWLLQDN 609
QY 525 AP-NSRYLSCPMESRHATYSNR-HKEN 550

RESULT 10
ID 013372 PRELIMINARY: PRT; 753 AA.
AC 013372;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN III FAMILY HOMOLOG.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 96230324.
RA XIANG R.H., HENSEL C.H., GARCIA D.K., CARLSON H.C., KOK K., DALY M.C.,
RA KERBACHER K., DEN BERG A., VELDHIJS P., BUYS C.H., NAYLOR S.L.;
RT Isolation of the human semaphorin III/F gene (SEMA3F) at chromosome
3p21, a region deleted in lung cancer.";
RL GENOMICS 32:39-48(1996).
DR EMBL: U38276; G1061351; -.
DR PFAM: PF00047; 19; 1.
SQ SEQUENCE 753 AA; 84941 MW; BECBBE80 CRC32;

Query Match 15.48; Score 688; DB 4; Length 753;
Best Local Similarity 31.48; Pred. No. 1,30e-141;
Matches 139; Conservative 121; Mismatches 140; Indels 42; Gaps 30;
Db 171 PYDPKUDTASALLNELLYAGVI-DEMGTDAAIFRLGKGTARTDOYNSRWLNDSEFIH 229
QY 139 PESPDSNLVLFEGDEYVSTIRKOEYNGK-IPRFRRIRGSELYTS--DIV-MQNQPFIK 194
Db 230 AEIIPSAENDKILFEFFERSAE-APQSPAVARYAGRIKLNDDGCHCL-VNKSFTFK 287
QY 195 ATIVHODQAYDDKIIYFFREDNDKNPAPLNVSRVAQLCRGDGOESSLSYSKNWTFEL 254
Db 288 ARLVCSVPGEGDGIETHFDELDVTV-OQTODV-RNPVYIAVFTSSGSGVFRGSAVVCYSMA 345
QY 255 AMLVCS--DAATNKNFNLODFLLPDPSCGWRDTRVGVSNPN-WNV--SAVCYSLG 308

[illegible]

Oy	524	LA-PNSRYLLSCMESRHATYSW	545
RESULT	12		
ID	Q9607;	PRELIMINARY:	PRT: 772 AA.
AC	Q9607;		
DT	01-NOV-1996 (TREMBLREL, 01, CREATED)		
DT	01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)		
DE	01-NOV-1998 (TREMBLREL, 08, LAST ANNOTATION UPDATE)		
OS	GALLUS GALLUS (CHICKEN).		
OC	EUAROTIA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;		
CC	NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-BRAIN:		
KX	MEDLINE; 94006554.		
RA	LDO Y., RAIBLE D., RAPER J.A.;		
RT	"Collapsin: a protein in brain that induces the collapse and		
RT	paralysis of neuronal growth cones."		
RL	CELL 75:217-227(1993).		
DR	EMBL; U02528; G410079; .		
SR	PFAM; PF00047; igf; 1		
SO	SEQUENCE 772 AA; 88867 MW; BC6CB9E8 CRC32;		
Query Match	15.3%; Score 681; DB 13; Length 772;		
Best Local Similarity	31.5%; Pred. No. 8,966-140;		
Matches	146; Conservative 119; Mismatches 150; Indels 48; Gaps 38		
Dd	166 RGKSPYDKLTSLVLVDGELVSGTA-ADEMGRDFAIFRTLGNHNPRTQNDHSRWLNDP	224	
Oy	135 RGVPFSPDENSLVLTFEGDEVSTIRKOENCK-IPIRRRIRGESELYTS-DTV-NQNP	190	
Dd	225 RFIAHLIPESDHPEDDDKIYFFERREALDGETLGATATARGIQCKNKDFEGHSL-VNKK	283	
Oy	191 QFIATIVHQ-DDAYDDKIYFFREDPNPKMEAPLNVSAQCICRDGGESSLSYSKW	249	
Dd	284 TTFKARLICSVGPNGIDTHPELDODVFLM-NSKDP-KNIYGVGTTSNIFKGAVC	341	
Oy	250 NTFLKMLAVS--DA-ATKNRNRIADVFLLPDPBGGMROTRVYGVSNNWN-Y--SAVC	303	
Dd	342 MYSTMIDRVAVELGYPVARNRGPNYOWPVYQGRVYPRPGTCSPTKFGFSFDTKDLPRVEVI	401	
Oy	304 VYSLGDIDKVERFS-SLK-G-YH--S--S-LPNRPCKCLPDQ-QEI-PLTEHF-QV-A	349	
Dd	402 FARSHPMYTPVPVINS-RPMIKTDVDYOQLIYVADVADSGO-YDVAFIGTDIGTYL	459	
Oy	350 -DR-HRPVAQRVRPMGLKTPLEFHSK--YHYOKAVAHMQASHETFNVLVTLTDROTII	405	
Dd	460 KVVISPKETHNELLEEVLLEEMTVFREPRYISAMKISTKOOOLYGSATGYSQLRHCYV	519	
Oy	406 KYVE-PEEOGHSAFNIME-IOPFRBAALIQTMISADARRKLIVSSQWEVSQVPLDICEY	463	
Dd	520 YGKACACECCLPDYCAWDSSGSSCRFYPRKARTKRODIRNGDLCTCSDLOHNHDPSGO	579	
Oy	464 YGGGCHGCCLMSRPHYCGMDGRCISITYSSE-RSYL-QSINAEPNHKPCPN-RPRDKAP--	518	
Dd	580 TLEEKIIYGVENSSTLECSPPYSQRAIYY-WOFQONKDNDHKE	621	
Oy	519 -LQ-KVSLA-PNSRYLLSC-PMESRHATYSWR-KRENVEDSGE	556	
RESULT	13		
ID	Q88632;	PRELIMINARY:	PRT: 785 AA.
AC	Q88632;		
DT	01-NOV-1998 (TREMBLREL, 08, CREATED)		
DT	01-NOV-1998 (TREMBLREL, 08, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBLREL, 08, LAST ANNOTATION UPDATE)		
DE	SEAPHORIN IV ISOFORM B.		
OS	MUS MUSCULUS (MOUSE).		
CC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA,		

M O S E R E
***** (TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:17:17 1999. Maspar time 29.68 Seconds

Tabular output not generated. 422.030 Million cell updates/sec

Title: >US-09-041-236-2
Description: (18-606) from US09041236.pep (32 of 45)
Perfect Score: 4401
Sequence: 1 SGPRIFAWKGVGHVGDVRDF.....SYFREAHQWQLPEDGIMAE 589

Scoring table:
PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseg35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.480; Variance 154.463; scale 0.236

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	661	15.0	771 13	R71380	Human semaphorin III	8.09e-51
2	620	14.1	477 13	R74175	Human collapsin.	7.99e-47
3	608	13.8	775 33	W63748	Human semaphorin.	1.17e-45
4	606	13.8	861 22	W17658	Mouse CD100 antigen.	1.84e-45
5	599	13.6	861 22	W58540	Human semaphorin.	8.78e-45
6	544	12.4	862 22	W17657	Human CD100 antigen.	1.86e-39
7	512	11.6	776 32	W51313	Rat semaphorin W.	2.28e-36
8	463	10.5	441 13	R71381	Vaccinia virus semaph	1.12e-31
9	459	10.4	730 13	R71379	Grasshopper semaphorin	2.69e-31
10	428	9.7	587 32	W51314	Human semaphorin W.	2.40e-28
11	390	8.9	974 33	W64221	Human secreted protei	9.43e-25
12	378	8.6	712 13	R71384	Tribolium semaphorin	1.27e-23
13	367	8.3	650 13	R71382	Drosophila semaphorin	1.36e-22
14	362	8.2	724 13	R71383	Drosophila semaphorin	4.05e-22
15	352	8.0	930 32	W57260	Human semaphorin Y.	3.45e-21
16	323	7.4	888 25	W19857	Human semaphorin Z.	1.11e-18

17	316	7.2	887 25	W19856	Rat semaphorin Z.	7.55e-18
18	296	6.7	929 32	W57259	Rat semaphorin Y.	5.20e-16
19	157	3.6	122 13	R71385	Varicella major virus s	8.38e-04
20	101	2.3	2329 25	W25038	Partial BRCA2 cancer	1.85e+01
21	99	2.2	591 4	R23006	Protein transcribed f	2.57e+01
22	98	2.2	832 33	W61092	Taq DNA polymerase I	3.03e+01
23	96	2.2	832 33	W61090	Taq DNA polymerase I	4.19e+01
24	91	2.1	478 1	R04881	Recombinant elastase.	9.30e+01
25	91	2.1	598 9	R48631	Sequence of nuclear r	9.30e+01
26	92	2.1	832 33	W61091	Taq DNA polymerase I	7.94e+01
27	92	2.1	854 30	W56309	Clas II S-receptor ki	7.94e+01
28	92	2.1	835 6	R29815	S receptor kinase pro	7.94e+01
29	93	2.1	4572 30	W52845	A. mediterranei rifam	6.78e+01
30	89	2.0	534 25	W25031	Partial BRCA2 cancer	1.27e+02
31	89	2.0	554 13	R66209	Novel thermostable DN	1.27e+02
32	90	2.0	638 5	R03924	E. coli HSP (dnak)	1.09e+02
33	89	2.0	680 4	R23143	Mutant thermostable D	1.27e+02
34	89	2.0	832 13	R76693	DNA-polymerase (F73L,	1.27e+02
35	89	2.0	832 13	P90556	Purified native therm	1.27e+02
36	89	2.0	832 33	W61087	Taq DNA polymerase I	1.27e+02
37	89	2.0	832 33	W61088	Taq DNA polymerase I	1.27e+02
38	89	2.0	832 33	R76690	Taq DNA polymerase RE	1.27e+02
39	89	2.0	832 33	W61089	Taq DNA polymerase I	1.27e+02
40	89	2.0	833 36	W59942	Amino acid sequence O	1.27e+02
41	89	2.0	833 27	W24211	Cleavase DA nuclease.	1.27e+02
42	89	2.0	833 27	W24212	Cleavase DA nuclease.	1.27e+02
43	90	2.0	1242 30	W52287	Rattus norvegicus cdo	1.09e+02
44	89	2.0	3418 26	W19211	Human breast cancer s	1.27e+02
45	89	2.0	3418 26	W23287	Human breast and ovar	1.27e+02

ALIGNMENTS

RESULT	ID	Query Match	Score	Length	DB
1	R71380	standard. Protein: 771 AA.	661	771	166 rtkxpydpkllasllldgelysgta-adfmgtrdfatrlghhprtqghdsrwindp 224
2	R71380	(first entry)	620	771	
3	21-NOV-1995		608	775	
4	Human semaphorin III protein.		606	861	
5	Human semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;		599	861	
6	Varicella major virus; smallpox; semaphorin receptor binding activity;		544	862	
7	modulation; nerve cell growth; immune response; viral pathogenesis;		512	776	
8	neurological disease; neuro-regeneration; oncological infection.		463	441	
9	OS Homo sapiens.		459	730	
10	PN W09507706-A.		428	587	
11	PD 23-MAR-1995.		390	974	
12	PF 13-SEP-1994: U10151.		378	712	
13	PA 13-SEP-1993: US-121713.		367	650	
14	PI (REGC) UNIV CALIFORNIA.		362	724	
15	PI Bentley DR, Goodman CS, Kolodkin AL, Mathes D;		352	930	
16	PI O'Connor T;		323	888	
17	DR WPI: 95-131177/17.				
18	DR N-PSDB: 087442.				
19	PT New class of semaphorin peptide(s) and polypeptide(s) - are				
20	PT potent modulators of nerve cell growth and regeneration				
21	PS Example 2; Page 60-63; 101pp; English.				
22	CC The sequence of the human semaphorin III protein. The proteins				
23	CC encoded by the grasshopper semaphorin I (087441), human semaphorin III,				
24	CC vaccinia virus semaphorin IV (087443), Drosophila semaphorin I and II				
25	CC (087444-5), Tribolium semaphorin I (087446) or varicella major (smallpox)				
26	CC virus semaphorin IV (087447) genes were used to generate a series of				
27	CC peptides. (R70370-R70418), which retain semaphorin receptor binding				
28	CC activity. The semaphorin derived or semaphorin receptor derived peptides				
29	CC are potent modulators of nerve cell growth, immune responsiveness and				
30	CC viral pathogenesis. They can be used in diagnosis and treatment of				
31	CC neurological disease and neuro-regeneration, immune modulation and				
32	CC diagnosis and treatment of viral and oncological infection and diseases.				
33	CC Sequence 771 AA;				
34	SQ				
35	Query Match	15.0%	Score 661	DB 13	Length 771
36	Best Local Similarity 30.3%		Pred. No. 8.09e-51		
37	Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;				

[illegible][illegible]

```

OY 198 VH0D0AY0DK1YYEFREDNPPKNEAPLANSVA0LCR00GEGSSLSJCKMNTFLKAML 257
Db 60 lcsvprpnqldthideldvflm-nfkdp-knpvvvgvfttsnifkgsavcmysmadvr 1177
OY 258 VCS--DA-ATKNKRNRLQDVEFLRDPDSG0M0DTRVYGVFNPNW-Y--SACVYSLDID 3111
Db 118 rvlfgpryhtrgprnyqvwpryvgtrvrprgtrcsrktctggtdsctkdldpdcvtatrspram 1777
OY 312 KVFRTS-SLK-G-YH-S-S-S-PNRPBPGCLSDQ-OPt-PEtP-F-OV-A-DR-NREV 3555
Db 178 ynpvfmnpnrplvltkdtnyqftglvvdrydaedqg-ydvmfigtvqvlkvsldpct 2361
OY 356 AQORVBP0KRLTP-L-FHSKYIYQKVAVNHNR0ASNGELFHNHLYLTIDGITHKVE-FCEQ 4133
Db 237 wydlaevlleemtyfcreptaisamejstckqqlyisgstagvaqlrhlrcdiygkacncc 2966
OY 414 EHSEFNFIME-IOEFPFRMAA10TSLDAERKRLTIVSSQWVSOVPLRDLCEVYGCGHCL 4722
Db 257 lardpawdsaesyifpctkxrttrtqdlngbrlthcsdlhbdhnhgspreoilliyv 3566
OY 473 MSRDYTCMD0GRCISYSS-BSVL-QSINPABNKECPKPKD---APLQKVSLA- 5255
Db 357 enstfliecsdpksqgrataylvwqfgrtneerkeelrvdhhllrtqglllra1qgkdsnyl 4166
OY 526 PMSRYLYSC-DMESRNATY-SW-NHKENV-EQSCPEHQ-SPNCLIEHNTLTAQOYCHNF 5800
Db 417 chavengftq 426
OY 581 CEADGSGYFR 590

```

RESULT 3
 ID W63748 standard: Protein: 775 AA.
 AC W63748;
 DT 01-OCT-1998 (first entry)
 DE Human semaphorin.
 KW Human; semaphorin; diagnosis; nerve disease; nerve growth inhibitor
 KW neurological disease; atopic skin inflammation; autoimmune disease;
 KW pain.
 OS Homo sapiens.
 PN W09822504-AL.
 PD 28-MAY-1998.
 PE 12-NOV-1997; J04111.
 PR 15-NOV-1996; JP-321068.
 PA (SUMU) SUMITOMO PHARM CO LTD.
 PI Furuyama T, Inagaki S.
 DR WPI: 98-312416/27.
 DR N-PSDB: V35367.
 PT Gene encoding new semaphorin nerve growth inhibitor - useful in
 PT diagnosis, treatment and study of neurological diseases
 PS Claim 1: Page 33-37; 49pp; Japanese.
 CC The present sequence represents human semaphorin, a nerve growth
 CC inhibitor. The semaphorin protein, and gene encoding the protein,
 CC and their derivatives, are used in the diagnosis, treatment and
 CC study of neurological disorders such as atopic skin inflammation,
 CC autoimmune diseases and pain.
 SO Sequence 775 AA;

Query Match	13.8%	Score 608:	DB 33:	Length 775:
Best Local Similarity	32.1%	Pred. No.	1,17e+45:	
Matches	125:	Conservative	98:	Mismatches 130; Indels 36; Gaps 28:
D b	167	rgrcpfcdpnssfvstlvgnelfaglys-dywgdrdsafirmsgkrlghrtchdderllkep	225	
O y	135	RGYAPFSDEDSILVLFEDDEVYSIRROEYNGK-IPFRRRI-R-GE-SELYTSSTVWNP	190	
D b	226	kfygsymipdhdnddkmkyffifekalaesenaehlytvtgrlcvdmwggqrll-vnkx	284	
O y	191	QEIKAIVHQDAVD-D-KIYYFFREDNDPKMEAPLVNRSVALCRLCDGCGESLSYSKW	249	
D b	285	stfikaarlvcsvpgmgidtyfdcaledvflrp-trdp-knpvfqglntsnslfrhnavc	342	
O y	250	NTFKLAMLVSCDAATN-KN-FNNLQVFLFLPDPDSGMGRJTRVYGVSNSPMN-Y-S-AVC	303	

Db 343 ylmmsiireanngprahkegpeymslsyegkvprpyrgscaskvngkkytctdypoda 402
 Oy 304 vylsldldikvrtts-slk-g---yhssl-----pnprrgcklpd-qop1--ptetf-qvad 350
 Db 403 rfarmrlpmyapikrvnhkpllvktcdgkynrlglavrgaeadg-ydvifigtdcivl 461
 Oy 351 rhpevaqav-epmcpk-ktf-lfms--kynhkaavhrrmoashgetrhvylttdkgtih 405
 Db 462 kvltlmgteetmmeevilleeqikfkdpailismelsiskrqvlgsasavayrflhcdm 521
 Oy 406 kvvergeqehsfafn-ime-iqprraaalqtsmldearrkrlvlyssqmevsyovpldcev 463
 Db 522 ygsaacdcclardrycawdgalscgrtpt 550
 Oy 464 yggcgchclmsrdpycgmdoarcslsyss 492

ID	RESULT	4
AC	W17658;	standard: Protein: 861 AA.
DT	24-JUL-1997	(first entry)
DE	Mouse CD100 antigen.	
KM	CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;	
OS	Mus sp.	
FT	Key	Location/Qualifiers
FT	peptide	1..41
FT	/label=	sig_peptide
FT	protein	42..861
FT	/label=	Mat_protein
FT	domain	42..553
FT	/label=	Semaphorin_domain
FT	domain	354..630
FT	/label=	Ig-like_domain
FT	domain	631..732
FT	/label=	Stalk_domain
FT	domain	734..752
FT	/label=	Transmembrane_domain
FT	domain	753..861
FT	/label=	Cytoplasmic_domain
FT	modified_site	807..814

FT /note- "putative tyrosine phosphorylation site"
 PN W09711366-A1.
 PD 15-MAY-1997.
 PT 12-NOV-1996; U18645.
 PR 09-NOV-1995; US-556422.
 PA (DAND) DANA FARBER CANCER INST.
 PI Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schulte JL,
 DR WPI: 97-780982/25.
 DR N-PSDB: T60666.
 PT Nucleic acid molecule encoding CD100 antigen - which stimulates
 PT leukocyte response, e.g. B cell aggregation, differentiation,
 PT survival and T cell proliferation
 PS Example 8: Page 86-89; 135pp: English.
 CS Mouse CD100 antigen (M17657) is a novel leukocyte semaphorin-like
 CC protein that stimulates a leukocyte response, including B cell
 CC aggregation, B cell differentiation, B cell survival and/or T cell
 CC proliferation. Its amino acid sequence was deduced from a cDNA
 CC clone (T60666) isolated from murine T cells. Human CD100 antigen
 CC (M17657) has also been identified. CD100 polypeptides and fusion
 CC proteins, nucleic acids, and host cells expressing CD100 can be
 CC utilised in diagnostic and therapeutic methods involving modulation
 CC of B and T cell responses, neuron axonal growth and immune cell-
 CC nerve cell interaction.
 SQ Sequence 861 AA;

	Query March	13.8%	Score 606;	DB 22;	Length 861;
	Best local similarity	32.4%	Pred. No.1,84=45;		
	Matches 145;	Conservative 100;	Mismatches 145;	Indels 47;	Gaps 31.
Db	107 eclnirv1qplststslvcvgctnaqfpqcdhlnltsfkllksedqkrcrfdpahsvts	166			

[illegible]

RESULT 5
ID W58540: standard; protein; 861 AA.
AC W58540:
DT 02-SEP-1998 (first entry)
DE Human semaphorin.
KW Human; semaphorin; diagnosis; nervous disease;
KW nerve extension inhibiting activity.
OS Homo sapiens.
PN J10155490-A.
PD 16-JUN-1998.
PE 27-NOV-1986; JP-332900.
PR 27-NOV-1986; JP-332900.
PX (SUMO) SOMITOMO SETYAKU KK.
PY WPI: 98-391044/34.
PZ DR N-PSDB; Y31121.
PT New human semaphorin gene - useful in the diagnosis of nervous system
PS and immune disorders
PS Claim 1: Page 10-12; 15pp; Japanese.
CC The present sequence represents human semaphorin (translated from the
CC Japanese specification as semaphorin). Semaphorin has nerve extension
CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
CC treatment and researches on nervous diseases and immune diseases.
SQ Sequence 861 AA;

	Query Match	13.6%	Score 599	DB 32	Length 861
	Best Local Similarity	32.4%	Pred. No. 8.78e-45		
	Matches 143	Conservative 103	Mismatches 146	Indels 47	Gaps 31
Db	107 eclnhvirvlgblststslvcgtaqpcldchlnltfkkfllgksedqgrcfcfpahyts	166			
QY	93 DCENITTLER-RSEGLLACTNMRHSC--WNLVNTVPLPGSM-RGYAPFSDEMSLV	148			
Db	167 vvmgagelysgts-yuflgsepibismnshsplrtetvayipwlnepsfvtadviqkspqge	225			
QY	149 LFEGGEVYSTIRKQENYNGKIPREFRIGSESLTSDTV--MNPQFIKATIVHQ--D-Q	202			
Db	226 gedcdvvyffitevsvevefvtklniprvaavcksdggqglitlg-kkwtstlkarloskp	284			
QY	203 AYDDKIYFFEDNDPKRPEAPLNVSRVAOLCRDQGESSLVSNNKTLKMLVCSDA	262			
Db	285 dsqivfnllqdvfnlrap-q1-keovfvavftoolnnvlsavcavylafveavfarskv	362			

QY 263 AATNKNFNRLQDVFLLPDPSCQWMDTRVYGVFSNPMWY---SACVSLGDIIDKVF-R--- 315
Db 343 msatvegshltkwryngpvrtpripacidsaaraanytslnlpdkltqfvkdhpmdd 402
QY 316 --TSSL-----K-G-YHSSLPNRPRKCL-PDQQPIR-TETFOVADHRPE-V-AQR-VE- 360
Db 403 svrpldnrpkllkdvnyqivvdrtqaldgtfydmfistdgaalhkavlltkvnh-v 460
QY 361 PMGPKLKT-P-LFHSKYNHOKVAVHRMQASHGETFHVLYLTDRGTHIKYVEPEQEHSEFA 418
Db 461 --leeqqlfrdfeprvltllsskkyrkfyagsgnsygvaplafeckhgs-ceedvlar 517
QY 419 ENIMEIQPFRRAAIIQTMSLDAER-RKL-YVSSQWESVQVPLDLCEVYGGCHGCLMSRD 476
Db 518 PYCAWSPAKACVTLHQEAS 538
QY 477 PYCGMDQG-R-CISITSSERS 495

RESULT 6
ID W17657 standard; Protein; 862 AA.
AC W17657:
DT 24-JUL-1997 (first entry)
DE Human CD100 antigen.
KM CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
OS Homo sapiens.
FH Key location/Qualifiers
FT peptide 1..41
FT /label_ Sig_peptide 42..862
FT /label_ Mat_protein 42..553
FT /label_ domain 554..630
FT /label_ Semaphorin_domain 631..733
FT /label_ Ig-like_domain 735..752
FT /label_ Stalk_domain 753..862
FT /label_ Transmembrane_domain 808..815
FT /label_ Cytoplasmic_domain 816..862
FT /label_ modified_site /note= "putative tyrosine phosphorylation site"
FT /label_ Phosphorylation /note= "putative tyrosine phosphorylation site"
PN WO9717368-A1.
PD 15-MAY-1997.
PR 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND) DANA FARBER CANCER INST.
PI Boussolet V, Freeman GJ, Hall KT, Nadler LM, Schultze JL,
DR WPI: 97-280982/25.
DR N-PSDB: T60665.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Claim 7: Page 70-72: 135pp: English.
CC Human CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60665) isolated by COS cell expression cloning. CD100
CC polypeptides and fusion proteins, nucleic acids, and host cells
CC expressing CD100 can be utilised in diagnostic and therapeutic
CC methods involving modulation of B and T cell responses, neuron
CC axonal growth and immune cell-nerve cell interaction.
SQ Sequence 862 AA;

Query Match 12.4%; Score 544; DB 22; Length 862;
Best Local Similarity 29.3%; Pred. No. 1,866-39;
Matches 157; Conservative 134; Mismatches 189; Indels 55; Gaps 39;

Db 107 eclnytrvrlqplslslyvcgtuafpacchlnltsfiflkknedgkyrcpfcphsyt 166
QY 93 DCENYITLLERSE-GLACGTNARHPSCMNLVMTVPLG--EM-RCYAFSPDEN-SL 147
Db 167 ymvvd-elysgts-yntfsgseplarsnshpdlteyaplpnepsfyfaivkspdsp 224
QY 148 VLFEDDEVYSTIRKOEYNGKIPFRFRIGSESELYTSDTV--MONDQFKATIVHO--D-- 201
Db 225 dgedrvyffftevseyefvrliprlarvcxqdgqglrtlg-kktsfikaarlcsr 283
QY 202 QAYDDKIYFFREDNPDKNPEAPLNVSRVADLCRDGEGSSLSYKNTFLKMLVCS 261
Db 284 pdsqglfvivldvrlzsp-gl-kxrvfyalfcpglnnvglavcaynlstaeevfshgk 341
QY 262 AATNKNFNRLQDVFLLPDPSCQWMDTRVYGVFSNPMWY---SACVSLGDIIDKVF--R 315
Db 342 ymgstvegshltkwryngpvrtpripacidsaaraanytslnlpdkltqfvkdhpmdd 401
QY 316 ---TS---S-LK-G-YHSSLPNRPRKCL-PDQQPIR-TETFOVADHRPE-V-AQR-VE 360
Db 402 dsvrpldnrpkllkdvnyqivvdrtqaldgtfydmfistdgaalhkai--s-lehav 458
QY 361 -PMGPKLKT-P-LFHSKYNHOKVAVHRMQASHGETFHVLYLTDRGTHIKYVEPEQEHSE 417
Db 459 hi-leeqqlfqqfepvqtlillskkyrkfyagsgnsygvaplafeckhgt-ceedvlar 516
QY 418 AFNIMEIQPFRRAAIIQTMSLDAER-RKL-YVSSQWESVQVPLDLCEVYGGCHGCLMSR 475
Db 517 dpycawspakacvtlhqeas 538
QY 476 DPYCGMDQ--GRCSITSSERSVLQSLNP-AEPHEKCPNRPDAKAPLOKVSALAPSRYLL 532
Db 575 kcsqgnlarvfwkfngvllkaespkyglmgknllf-nlsegdsgygcglsee 628
QY 533 SCPMSRRIATYSMRKKEVNEQSCP--GHQSPNCLIFLENTTAQYXGHPCEAD 585

RESULT 7
ID W51313 standard; Protein; 776 AA.
AC W51313:
DT 08-SEP-1998 (first entry)
DE Rat semaphorin W.
KM Rat; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KM immunosuppressant; gene therapy; diagnosis; research reagent.
OS Rattus norvegicus.
PN WO9815628-A1.
PD 16-APR-1998.
PR 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T,
DR WPI: 98-261015/23.
DR N-PSDB: V07279, V07280.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Claim 1: Page 60-64; 90pp: Japanese.
CC The present sequence represents rat semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 776 AA;

Query Match 11.6%; Score 512; DB 32; Length 776;
Best Local Similarity 27.1%; Pred. No. 2,256-36;
Matches 152; Conservative 137; Mismatches 218; Indels 53; Gaps 37;
Db 61 rfaashlynygallyvdpsashlyvgardsfalclpfsgerpridwmp-ethrncrk 119
QY 34 RVDGQTEPHVTLFHEPSSSVWVGGRKYV-L-FDEPEGNASV-RVYNIGSTKSGCL- 89

RESULT

ID R71381 standard; Protein: 441 AA.

AC R71381;

Dt 21-NOV-1995 (first entry)

DE Vaccinia virus semaphorin IV protein.

KM Vaccinia virus; grasshopper; human; vaccinia virus; Drosophila; Tribolium;

KW variola major virus; smallpox; semaphorin receptor binding activity;

KX modulation; nerve cell growth; immune response; viral pathogenesis;

KN neurological disease; neuro-regeneration; oncological infection.

OS vaccinia virus.

PN MO9507706-A.

PD 23-MAR-1995.

PE 13-SEP-1994; U10151.

PR 13-SEP-1993; US-121713.

PA (REGC) UNIV CALIFORNIA.

PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;

I O'Connor T;

DR MPI: 95-131177/17.

N-PDB: 087443.

PT New class of semaphorin peptide(s) and polypeptide(s) - are

PS potent modulators of nerve cell growth and regeneration

Example 2: Page 65-67; 101P; English.

The sequence of the vaccinia virus semaphorin IV protein.

The gene sequence was isolated as the A39R open reading frame sequence

from variola, based on sequence homology searches of a database with the

grasshopper, Tribolium and Drosophila semaphorin sequences. The proteins

encoded by the grasshopper semaphorin I (Q87441), human semaphorin III

(Q87442), vaccinia virus semaphorin IV, Drosophila semaphorin I and II

(Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)

virus semaphorin IV (Q87447) genes were used to generate a series of

peptides (R70370-R70418), which retain semaphorin receptor binding

activity. The semaphorin derived or semaphorin receptor derived peptides

are potent modulators of nerve cell growth, immune responsiveness and

viral pathogenesis. They can be used in diagnosis and treatment of

neurological disease and neuro-regeneration, immune modulation and

Cc

Dd 120 kggkdeechffigllaivnaashlltcgtafadpkcgivdfvssfgqverlesgqrkcpcfe 179

90 -DKK-D-CENTITLLERSEG-LIACGTAAHSPSCNNL-VNG--TWPLGEMRGAPFSP 142

Dd 180 agraaamagvglyatavk-nflgtepelisavgradevlretlisswlnapaayaml 238

Oy 143 DENSLVFEGDEVSTTRKOETNGKLPFRIRIGESSELYT-SDTV---MONPOFKATIV 198

Dd 239 spaewgedddeliffittetsrvldsyerikvyryarvacagdligrktlg-qwtetfkl 297

Oy 199 -H-Q-DQAVDDKIYFEFREDNDPKNDEAPLANVSFAVAQCRRDGGESSLSVKNNFTLK 254

Dd 298 adllcppehngirasygvqlamaelrpqpg-a-gdrilfigtsqweagaalsavcafrpdqr 356

Oy 255 AMLCSSAATANKNFNRQADVFLLRPDSGGWRDRYRVGVSNWN--YSACVVSYLGID 311

Dd 357 avlmgpfrelkhdcnrglpymdhvepqprgecianmmlqgfsgslspdtclffirdh 416

Oy 312 KVFRTS--SLK-GHSFLP---N-----PRGRKLPPQQOIP--TEFFVADNHPE-V-AQ 357

Dd 417 plmdrpfpadgrrpllvttctlaylrvaahvtelsgkeyviyligtcdghlhavr1gaq 476

Oy 358 R-VE-PGMPKLT-PLFSHK-YHYOKAAVHRMQASHGEFHVLVLTLDROGTHKVPEDQ 413

Dd 477 -ls-vldedlal-fpeqpvesmklyhdw-llvgshtevtyntsngrlgs-csecll 529

Oy 414 EHSAEAFIMEIQPRRAAIQTMSLDERKKLVSSQSWESQVRDLDCVYGGGHGLM 473

Dd 530 agdvncasfrldacvchagehgrgmvgdielasvaslcckepgehprvteyvatacvghv 589

Oy 474 SRDPYCGWD-Q-GRCISIVSESRVLSQINAPAHKECCNPDPKAPLCKVSLAANSRY 531

Dd 590 lpcpsasaascvwbpsy 609

Oy 532 LSCPMESRHATYSMRKENY 551

CC	diagnosis and treatment of viral and oncological infection and diseases.
CC	Sequence 441 AA.
Query Match	10.5%; Score 463; DB 13; Length 441;
Best Local Similarity	33.9%; Pred. No. 1.12e-31;
Matches	87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;
Db	115 lvgctungnpkckw-ldgsddphkrgyapygnskvtlshnmc-vlsdlnisk-eg-1 170
OY	109 LACSTNARHSPSCMNVLVGVTPVGEGRGAYAPSPDENSLVFEGDEYSTRKOEYNGKI 168
Db	171 kwrtfdgpcgyllycnadnvpkdg-lygafcdkgctyd-kvylftctlskr--1-vk 225
OY	169 PRRFRIGE-S-LLYTSDVWQNPQFIKTYIHQOQAYDDKIKYFFERDNDPKNPEAPLN 226
Db	226 lpyhagnclddegppsslshtwstflkxvelcc-did-grey-r-q-lhsrtlkt-d 279
OY	227 VSRPAQLCRDQCGESSLSVKRNITLKMALVCSDAIANKNPNRQLQDYFLLPDSGQMRD 286
Db	280 tllvffidspysksaalcymsmltkqsfstsklegtykqlpspaaglcldpgkxvphltf 339
OY	287 TRVGVFSPNPMNYSACVYSLGIDIKVFRFTSLKGYHSSLPBRPGKCLPQOQIPRETF 346
Db	340 evlekyvlddlkpkis 356
OY	347 QVADRPHEVAQRAPEPMG 363
RESULT	9
ID	R71379 standard: Protein: 730 AA.
AC	R71379;
DT	21-NOV-1995 (first entry)
DE	Grasshopper semaphorin I protein.
KM	Semaphorin: grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW	varicella major virus; smallpox; semaphorin receptor binding activity;
KW	modulation; nerve cell growth; immune response; viral pathogenesis;
KW	neurological disease; neuro-regeneration; oncological infection.
OS	Grasshopper sp.
PN	W09507706-A.
PD	23-MAR-1995.
PF	13-SEP-1994; U10151.
PR	13-SEP-1993; US-121713.
PA	(REGC) UNIV CALIFORNIA.
PI	Bentley DR, Goodman CS, Kolodkin AL, Mattes D;
PI	O'Connor T;
DR	WPI: 95-131177/17.
DR	N-PSDB: 087441.
PT	New class of semaphorin peptide(s) and polypeptide(s) - are
PT	potent modulators of nerve cell growth and regeneration
PS	Example 1: Page 68-72; 10pp; English.
CC	The sequence of the grasshopper semaphorin I protein. The proteins
CC	encoded by the grasshopper semaphorin I, human semaphorin III (087442),
CC	vaccinia virus semaphorin IV (087443), Drosophila semaphorin I and II
CC	(087444-5), Tribolium semaphorin I (087446) or varicella major (smallpox)
CC	virus semaphorin IV (087447) genes were used to generate a series of
CC	peptides (R70370-R70418), which retain semaphorin receptor binding
CC	activity. The semaphorin derived or semaphorin receptor derived peptides
CC	are potent modulators of nerve cell growth, immune responsiveness and
CC	viral pathogenesis. They can be used in diagnosis and treatment of
CC	neurological disease and neuro-regeneration, immune modulation and
CC	diagnosis and treatment of viral and oncological infection and diseases.
Sequence	730 AA:
Query Match	10.4%; Score 459; DB 13; Length 730;
Best Local Similarity	28.0%; Pred. No. 2.69e-31;
Matches	143; Conservative 130; Mismatches 182; Indels 56; Gaps 46;
Db	32 qfgeeyqrflgneskhfkllkdxnsllygarniivynslrdlftfegqrlwshsg 91
OY	29 HVGQDRND-FGQREPIITVLFH--EFGSSSVWVGKGKRYLFDPEGKN-ASVRIIV-NI-G 82
Db	92 ahelecylykxskdsdcqnylrvlakiddrvllcgtlnaykplcrhyalxkdgdyvvekeye 151

QY 83 STKSC-L-DKR--DCENVITLLERRSEG-LLACGTNARHPSGMN--LVNCT-VVPLG-E 133
Db 152 grglcplcdpnhnstaiaiyseaglysatv-afdsgrcdp-11-yrg-p-ltersdl-k-q1n 204
QY 134 MRGAPRSPDENSILVFECDDEVSTIRKOEYNGKIPRRRRIRGESELYTSPTVMONPOFI 193
Db 205 apufvn-tmeyndfiffifretaveylnockaiy-srvarvckhdkgg-phgggdrtwtsf 261
QY 194 KATIVHODQAVDDKIYFFREDNDP-KNPEAPLWNSVAQLCRGDOGESSLSVSKWMTF 252
Db 262 lksrlncsvpgdyfifvfeistcdillegnygg-vekllygvftlpvnsygsaavcafs 320
QY 253 LKAALVQSDAATNK-NFNRLODVF-LLEPD-SGQWRDRYGVFSNPMN-Y-SAVCYSS 306
Db 321 mksllsfdfgfkgeqetnsmwlvapsikvpeppgcvnscrtlpdsvsnfvskhltmd 380
QY 307 L-G--D-IDKVFRT-SLTKG-YHS--SL--PNPRPGKLPDQOQPIPTTFQVADRHPEVA 356
Db 381 gawpafit-epillrtislyqrfcklavdqvyrtpdgkaydlf1gtdgkvika1nsesf 439
QY 357 QRVPMGRLKTPLEH-S-KYHOKVAV-HRMQASHGETFHFVLYLTDRGTHKVEPG-- 411
Db 440 dsastvsvleelqvpvpyvknlyvrmdgdsdklvvssdellaiklhrgscklt 499
QY 412 EQEHSFAFNME-IQPFRRALIQTM--SLDAERKRLYSSQWESOVPLDLCEVYG-G 466
Db 500 ncrecva1qdpqcawvnelkctavspdw 530
QY 467 GCHGCLMSRDYPCGMDQ-G-RCISYSSERS 495

RESULT 10
ID W51314 standard; Protein: 587 AA.
AC W51314;
DE 08-SEP-1998 (first entry)
KW Human semaphorin W.
KW Human semaphorin W; nerve extension inhibitor; anti-allergic; anticancer;
KW Immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PN MO9815628-A1.
PF 16-APR-1998.
PR 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR MPI: 98-261015/23.
DR N-PSDB: V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC anti-allergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 9.7%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 2,40e-28;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgddeliffetstafdsyerikprvavcagdlggrkltq-grwtftladl1p 115
QY 201 DQAVDDKIYFFREDNDKNEAPLWNSVAQLCRGDOGESSLSVSKWMTFLKAMLVCS 260
Db 116 gphgrssv1qdvavlrpelga-gtbflyg1fssqvegatsavcafrpdditv1npg 174
QY 261 DAATNKNFNRRQDVFLLPDPSCGWRDRYGVFSNPMN---YSAVCYSLGIDIKVFRIS 317
Db 175 frelkhcengrlpvvndvdpqprpgecltnmklrhfgsslsldrvl1tflrdhplmdrp 234

QY 318 --SLK-GYHSSLP---N-----PRGKCLPDQOPIP--TETFFQVADRHPE-V-AQR-VE-P 361
Db 235 vfpedgphllvttdctaylvrvahrvtslsqkeyvlylgdedghlhnvav1gaq-1s-v1 292
QY 362 MGPLKT-PLFHSK-YH1QKAVHRMQASHGETFHFVLYLTDRGTHKVEPGQESHFAF 419
Db 293 edlal--fpepgrvemk1-yhs-w-1lvsrtevtqyntcgrlgs-csecllagdpvc 347
QY 420 NIMEIQPFRRALAIQMSLDAERKRLYSSQWESOVPLDLCEVYGCGCHCLMSRDPYC 479
Db 348 awsf1lecvaahgehrglvqdlasvslcpkegeprvrvfepvataahvlp1pcsp 407
QY 480 GWD-Q-GRCISYSSERSVQSLINPAEPHKECPNPKDKAPLQKVS1APNSRYLSCPME 537
Db 408 sawscvwhpsgy 421
QY 538 SRHATYSWRHKNV 551

RESULT 11
ID W64221 standard; Protein: 974 AA.
AC W64221;
DE 06-OCT-1998 (first entry)
KW Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokine; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN MO9827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-768192.
PR 13-JAN-1997; US-763401.
PA (GEMV) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR MPI: 98-362774/31.
DR N-PSDB: V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokine activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 8.9%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 4,43e-25;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechmfikylknddalvfcgtnafnpscrnykndtlepfdfetsgmrcpydakhan 172
QY 91 KRDCENYIT-LLERRSSGLACGTNARHPSGMN-LVNGTVPLG-EKRGVA--PFSPDENS 146
Db 173 valfaedklysatv-dflaidavlyrs1-gesp--tlrtvkhskwlykeryfvagvdyg 228
QY 147 LVTFEGDEVSTIRKOEYNG-KIPRRRIIRGESELYTSPTVMONPOFIKATIVHODQAVD 205
Db 229 dylyffirelavenyngkvfrvavcncndmgsgqrvlekwtsf1kar1ncsvpgds 288
QY 206 DKIIYFFREDNDKNEAPLWNSVAQLCRGDOGESSLSVSKWMTFLKAMLVCSDAATN 265

Db 289 hffnllldavcvtvrling--tdv-vlfttstrynslbrgsavcauamldlavftgfrkq 345
Oy 266 k-nferrlodvllrddpsqwmtdtvtvufvfnspmh-y--sawcyvsgldidkve--rtssl 319
Db 346 kspdstvrvpdelevrpkprccagssslrtaetnaefddtlnflktrplmdaavrsif 405
Oy 320 kgynhs---lnrprgkslrr--qqrlptetrvoadnhrva-q--rver-m--g-p-l- 365
Db 406 nprwfltmvtyrltklavdaagry-qnhltvufldgeakllkflarignsgflndslf 464
Oy 366 ktprle-hs--kyunokvaunhmqasohetgrhnyulttdrctikrver--geof--hsfaf 419
Db 465 leemsvynseksgdyvedkrilmqtdressalyaafscvklrvpgrcethrg-cktt 523
Oy 420 -nimef-qprfra-aaio---tksldbr--krllyssomevqvrpdlcsevggch-g 470
Db 524 clasdrvcywlkeggagacshlspnrlltfcdqlern 560
Oy 471 clmsdrpycgw--dogrcisylssersvul--osinpa 504

RESULT 12
ID R71384 standard; Protein: 712 AA.
AC R71384;
DT 21-NOV-1995 (first entry)
DE Tribolium semaphorin I protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Tribolium sp.
PN W09507706-A..
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T.
DR WPI: 95-131177/17.
DR N-PSDB: Q87446.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2: Page 85-88; 101pp; English.
CC The sequence of the beetle Tribolium semaphorin I protein. The gene was
CC isolated by PCR using Tribolium genomic DNA. The proteins encoded by the
CC grasshopper semaphorin I (Q87441), human semaphorin IIR (Q87442),
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-45), Tribolium semaphorin I or variola major (smallpox) virus
CC semaphorin IV (Q87447) genes were used to generate a series of peptides
CC (R70370-R70418), which retain semaphorin receptor binding activity. The
CC semaphorin derived or semaphorin receptor derived peptides are potent
CC modulators of nerve cell growth, immune responsiveness and viral
CC pathogenesis. They can be used in diagnosis and treatment of neurological
CC disease and neuro-regeneration, immune modulation and diagnosis and
CC treatment of viral and oncological infection and diseases.
SQ Sequence 712 AA;

	Query Match	8.68;	Score 378;	DB 13;	Length 712;
	Best Local Similarity	27.04;	Pred. No. 1.27e-23;		
	Matches 140;	Conservative 128;	Mismatches 194;	Indels 56;	Gaps 43;
Db	48	hfiyinq-detsilvggrmrvynslsifdsergsgridwps-sdahqglclkgtdcdc	105		
	: : : : : : : : : : : : : :				
OY	43	hVVLHREPGSSSVWVGGRKV-L--FDPEEGKNASRYTINISTKGS-C-L-DKRQ-C	94		
Db	106	guyirilyssepqklvcgctnsykplicrtyafekqylvekevegiglcypnepnhsstsv	165		
	: : : : : : : : : : : : : :				
OY	95	ENYITDLERSES-LIACGIRNARHPSCWNLV--NGT-VVFLG-EMRGTARFPDENSL-V	148		
Db	166	syng-qfisaetv-adfsggdpllyreptetl-sdlkqlnapnfvs--v---aygdyi	216		
	: : : : : : : : : : : : : :				
OY	149	LEEGEVEVSTIRKOENNGKIPRRRIIRGESEELVTSDFVOMNPGIKTIYHODADYDKI	208		

Db	217	ffiyetaveymcgkviiy-svataevckdkkg-phgsidevtsfllkarlncsipyepf	214
Oy	209	yyffefrednpd-knpeaplnvsvaolcsgdggesslsvskmftlamlvcspdaatnk-	266
Db	275	yfdelqgsdivegrynsdskliiygllttrpnaigssatcayqmedllrvfegstkq	334
Oy	267	nfnrlodvflldpdsqowdrv-vygfnsnmv-y-saacvylslcdldkvf---R-T	316
Db	335	etlmsnwjpyvqnlvpepripqcvrdrslldkvnflkhtslmed-vpalfg-kpvlvr	392
Oy	317	sslkv-y-h-s-sl-pndrpeclldpddpirtetffoyadhrhveayorvberpuktlfn	371
Db	393	vslygrfcaitevpdqvtlknqyldvlylgrtdgkvllkavnlpkrthakallykyrtsvh	452
Oy	372	-s-khvgkvaun-rmqashetfnhlyltldrgrtlnkve-peoghsfapnmetlopf	427
Db	453	phgarvklkllarygvkvvvvgkdeitlanlnhas-ktvcddvevlqdfhcawdaqn1	511
Oy	428	rraaaiqtmslmdaerkrklyvssosvevsopvldlcevyvgcgchgcslmsrdpycwmD--QGR	485
Db	512	csvidtvsyflilqdvvggdgn-cwspqdkktvix	548
Oy	486	cistys-se-rsvlqsinpaprnhceprkdkpdlrk	521

RESULT 13
ID R71382 standard; Protein: 650 AA.
AC R71382;
DT 21-NOV-1995 (first entry)
DE Drosophila semaphorin I protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW varitola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Drosophila sp.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; 010151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
PI WPI: 95-131177/17.
DR N-FSDB; 087444.
DR New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 74-77; 101pg. English.
CC The sequence of the Drosophila semaphorin I protein. The gene was
CC isolated by PCR using primers based on sequence homology between the
CC complete grasshopper and partial Tribolium and Manduca (moth) semaphorin
CC gene sequences. The products gave two different sequences, each of which
CC was used to obtain its respective complete sequence: semaphorin I and II
CC (087445). The proteins encoded by the grasshopper semaphorin I (087441),
CC human semaphorin III (087442), vaccinia virus semaphorin IV, Drosophila
CC semaphorin I and II, Tribolium semaphorin I (087446) or varitola major
CC (smallpox) virus semaphorin IV (087447) genes were used to generate a
CC series of peptides (R70370-R70418), which retain semaphorin receptor
CC binding activity. The semaphorin derived or semaphorin receptor derived
CC peptides are potent modulators of nerve cell growth, immune
CC responsiveness and viral pathogenesis. They can be used in diagnosis and
CC treatment of neurological disease and neuro-regeneration, immune
CC modulation and diagnosis and treatment of viral and oncological infection
CC and diseases.
SQ Sequence 650 AA;

	Query Match	8.3%	Score 367;	DB 13;	Length 650;
	Similarity	28.4%	Pred. NO. 1,35e-22;		
	Matches	124;	Mismatches 158;	Indels 51;	Gaps 38;
	Conservative				
Db	3 degnylrmvpspgirifvcgtsfrmcmtlylsdcsnytleatknqgavcpvdpnrnst 62				
ov	93 DCENLTTLERRSG-ILACGTNRHRSCKNW-IVNCGVVLGEMR-GYA--PSPSPNST.147				

[illegible]

RESULT 14
ID R71383 standard: Protein: 724 AA.
AC R71383;
DT 21-NOV-1995 (first entry)
DE Drosophila semaphorin II protein.
KW Semaphorin: grasshopper; human: vaccinia virus: Drosophila; Tribolium;
KW varicella major virus, smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Drosophila sp.
PN W0507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (RGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB: 087445.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2: Page 79-82; 101pp: English.
CC The sequence of the Drosophila semaphorin II protein. The gene was
CC isolated by PCR using primers based on sequence homology between the
CC complete grasshopper and partial Tribolium and Manduca (moth) semaphorin
CC gene sequences. The products gave two different sequences, each of which
CC was used to obtain its respective complete sequence: semaphorin I
CC (087444) and II. The proteins encoded by the grasshopper semaphorin I
CC (087441), human semaphorin III (087442), vaccinia virus semaphorin IV,
CC Drosophila semaphorin I and II, Tribolium semaphorin I (087446) or
CC varicella major (smallpox) virus semaphorin IV (087447) genes were used to
CC generate a series of peptides (R70370-R70415), which retain semaphorin
CC receptor binding activity. The semaphorin derived or semaphorin receptor
CC derived peptides are potent modulators of nerve cell growth, immune
CC responsiveness and viral pathogenesis. They can be used in diagnosis and
CC treatment of neurological disease and neuro-regeneration, immune
CC modulation and diagnosis and treatment of viral and oncological infection
CC and diseases.
SQ Sequence 724 AA:

Dd	265	vyffifretaeavyincogkavy-sriarvckckdvvgkknla-hmavtykaltncisgcfp	322
Oy	208	tyfferebnnp-knppearlwnsryvaolcrrdgoggeslsvskmwteflkaalwcsdaatnk	266
Dd	323	fyfneigsavvylspdkarffat--ftlstnglisyavcsfhlnelgaafngkfkqessn	380
Oy	267	-nwnrlodvelflppdpsgwmrtvryugvfnspwnsnavcyytsgldid--kv-fr--ts-	317
Dd	361	saaylrvlnsryperptcyvndtsnrbctvlnftrshplmdkavyn-eh-npnyuykrdl	438
Oy	318	s--lkgvhsslpnhrpfcscldoqprlctetfvoaadrhpvaorvbrpcklptlfhshk-y	374
Dd	439	vftklnvdkilridlungeylyuyvngtlngrfyklyvuyngests-klidfevareat	497
Oy	375	hkyovavahrmqas-hgefnhlyllyttrgrtlnhkvvegeedehsrafimltioframaat	433
Dd	498	qymetsqtklslylgtchdnhkqidlamcnrtgdn-cftc-v-trdyrcswd	544
Oy	434	qtmstldearrklyvssqmwesovpdlcsev-yggscgncslmsrbrpysgmd	482

RESULT 15
ID W57260 standard; Protein; 930 AA.
AC W57260;
DT 02-SEP-1998 (first entry)
DE Human semaphorin Y.
KW Human, semaphorin Y; nerve extension inhibitor; central nervous system;
KW peripheral nerve growth.
OS Homo sapiens.
PN W09811216-A1.
PD 19-MAR-1998.
PF 09-SEP-1997; J03167.
PR 08-AUG-1997; JP-227220.
PR 11-SEP-1996; JP-263565.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-250958/22.
DR N-PSDB: V28915, V28916.
PT DNA encoding human and rat semaphorin Y - an inhibitor of nerve
PT extension
PS Claim 1; Page 65-70; 85pp; Japanese.
CC The present sequence represents human semaphorin Y, which inhibits
CC nerve extension. Semaphorin Y genes and proteins may be used to inhibit
CC peripheral nerve growth. Semaphorin Y antagonists can be used to
CC accelerate regrowth of the central nervous system.
SQ Sequence 930 AA;

[illegible]

OY 367 -TPEFH--SKHYQKVAHHRMQASHGETHVLYLTDRGITHKVEPEGEOHSEFNFIME 423
Db 470 eidaysparcsgkrtatarrilgleldteghrlfvafigclvy1p1srcarhga-cqrs 528
OY 424 -IQPF---R---RAA--A--IOTMSLDAERKRKLIVSSQWENSOVPLDLCFVYGGGCH-G 470
Db 529 clasqdpqcgwhssrgcvtlrgsgtldvdqagngesmehdcdq 572
OY 471 CLMSRDPYCGWDOGR-CISIVSSE-RSVLQISINPAE-PHKECPN 511

Search completed: Thu Jul 8 19:17:53 1999
Job time : 36 secs.

THIS PAGE BLANK (USPTO)

Qy	378	KVAHNRQASNGEFTF-VLLITDRGIHKKVBERGEDESHSFANNIMEIQDFRRAAIQTM	436
Db	491	LDDNTNLKLVYSEWESEVPLDCSYGNDCEFCSEMSRDPCLCTWYNNTC-S-FK-QRVS	547
Qy	437	SLDERRLKLYSSQWESVQPLDCEYVGGCGGCGICLMSRDPYCGMOCGICISYSSERSV	496
Db	548	VETGGPARNRTISEMGCGHYPATVYKHQVSTPLNSNLSLCPANSHADFMFKDGETER	607
Qy	497	LOSINPAEPH-KE-CPNPKPDKAPLQVSLAPNSRYILSCPMESRHATYSWRHKEVENS	554
Db	608	CHVTHKNDCLILIANSTATNGTHVCNMKEDS	640
Qy	555	CEPGHSPNCILFIENLTAAQYGHYFCEAEGS	587
RESULT	2		
ENTRY	148744	#type complete	
TITLE	semaphorin A - mouse		
ORGANISM	#formal_name Mus musculus	#common_name house mouse	
DATE	02-Jul-1996	#sequence_revision 02-Jul-1996	#text_change 04-Sep-1998
ACCESSIONS	148744		
REFERENCE	148744		
#authors	Puschel, A.W.; Adams, R.H.; Betz, H.		
#journal	Neuron (1995) 14:941-948		
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension		
#cross-references	MIMD:95267431		
#accession	I48744		
#status	preliminary: translated from GB/EMBL/DBJ		
#molecule_type	mRNA		
#residues	1-748	#label RES	
#cross-references	EMBL:X85990; NID:9854323; PID:9854324		
GENETICS			
#gene	sema		
CLASSIFICATION	#superfamily semaphorin		
SUMMARY	#length 748	#molecular-weight 82894	#checksum 9017
Query Match	16.7%	Score 737; DB 2; Length 748;	
Best Local Similarity	33.7%	Pred. No. 1,45e-134;	
Matches	168;	Conservative 112; Mismatches 166; Indels 52; Gaps 36;	
Db	112	ECANFVLLAHYNTLLACRTGAFHTCALRMATAGGTHASTGPEKLEDGKGTYPD	171
Qy	93	DCENVITLLERSSCG-LIACGTNARHPSG---WNLVNGT-V-V-P-LGEMGVAPESP	142
Db	172	RHRPVSUVGEELISGV-TADLMGRDPTTIRSGONPSLTBEHDSRMLNPEFYVFWI	230
Qy	143	DENVLVEFGEDEVSTIRKQDYNGK-IPRRRTIRGESELYT-S-DIV-MONPOFIAT-I	197
Db	231	PESENPDDDKIYFFRESAVEAAPAMGRMSVSRYGICRNDLGORSI-VNKKWTFLLKAR	289
Qy	198	VHQQAADDKIYFFREEDNDPKNE-APLWVSRAQLCRDQGESSELSVSKWTFLLKAM	256
Db	290	LYCSVPEVEGDTHDOQDYFELLS-SR-DQOTFLAVFTSSGVSQGSACVYSNDVXR	347
Qy	257	LYCSV-DAATKKNRNRLQDVFLLPDSGQRODRTYGVGSNPNV-Y-SAYCVYSIGDID	311
Db	348	RAFGPRLPHKEGPHQWVSQGRAPYRPMCSKTFGFSSTKDPDDVDIOFGRRHPLM	407
Qy	312	KVFETIS-SLK-G-----YHSSLPNRPRCKLPDQ-QRIP-TEFP-QVAD-R-HPEV	355
Db	408	YNVPLPMGQ-R-PLFLOVGAGYTFQTQIADRAVAADGH-YDLFICTDVGVTLKVISVRK	464
Qy	356	AQRVEPMGPKLTPF-H-SKYNHOKAAVVRMOASHGETHFVHLYTTDGTGTHIKVPE-P-	410
Db	465	GRPNSSGGLLEELQVEDSAITSMQISSKROOLVVASAAVAQIALRCAALBAGACAE	524
Qy	411	GEQHSSTAFNIMEIQDFRRAAIQTMSLDERRLKLYSSQWESVQVPLDCEYVGGCGHG	470
Db	525	CCLARDYGCAMDGSACTRFQPTARRRRRDDINDGSPSTLCSGDSHSLVLEKKVLGVES	584

```

OY 471 CLMSDPRVCWGDQRCISYSSESYSVL--QSINPAERHKRCPRKDDKAPLQVSLAPNS 528
Db 585 GSATLECEPRSLQAHVQW 602
OY 529 -RYLLSCPMESRHAITYSW 545

RESULT 3
ENTRY 148748 #type complete
TITLE semaphorin E - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998

ACCESSIONS
REFERENCE 148748
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
family and creates domains inhibitory for axonal extension
#cross_references MVID:95267431
#accession 148748
#status preliminary: translated from GB/EMBL/DBD
#molecule_type mRNA
#residues 1-751 #label RES
#cross_references EMBL:X85994; NID:9854331; PID:9854332
GENETICS
#gene semaphorin
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 751 #molecular_weight 85259 #checksum 8961

Query Match 16.5%; Score 725; DB 2; Length 751;
Best Local Similarity 33.0%; Pred. No. 8,46e-132;
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;

Db 167 SFNPVNVVSMINDELFSQNYI-DEGTLAALFRSLTKMOLRTDOHNSKMLSEPMFVD 225
OY 139 PPSDENSELVLESGDEYVSTIRKQETNGK-IPRRRIKGESELYTS--DTV-MQNQPIK 194
Db 226 AHVIPDGDTPDAVYVEFFEKRLTDNNRSTKQIHSIARIKCPMDTGGORSI--VNKWTFL 284
OY 195 ATIVHQD-QANDDKIYFFREDNDKRPAPLAVSRKLAQCRDQGESLSKNTFL 253
Db 285 KARLYCSTDEGDEPHDELDEYFL-ETDNP-RITLVYGITTTSSVFKGSAVCYHL 342
OY 254 KAMLVCS--DA-ANKNENRLODFVFLPDSGQWRDTRVYGVSPNSWN-Y--SAVCYSL 307
Db 343 SDIQTVENGPRAHKEGPHOLISVQGIPIPRPRTGCGAFTPNMRTTDFDPDVTFR 402
OY 308 GDIDVFTS-SLR-G-----YHSLPMPRREKCLIPDQ-QP-I-PTLTF-QVADRRP 353
Db 403 NHPLMYISIPRIHRPLIVRTGIDYKTKIAYDRVNAADG-RYHVLFLGDRGTQKVVV 461
OY 354 EVAQREVEPMGSL-KTPIF-H--SKYHQKVAVHRMQASHGEPHYLVLTLDROTIIKVE 409
Db 462 LPTNSASGELLLELEVFKNHPVITTMELSSKKOOLYSSNNGVQVSLHRHITGTC 521
OY 410 -PGDEHSAFNINELIOPFRRAALIQTMSDAERRKLYSSQWVEVQVPLDCEYVGGGC 468
Db 522 ADCCLAPDPYCAMGHCSCRFYPGKRKRSRROVRHGNPLTQCRGNLKAAYNAAEIVQY 581
OY 469 HGCLMSRPYCGMQGRCISITYSSE-RVSLQGINPAERHKCP--NPKRDKAPLQKVEL 524
Db 582 GVRNNTFLLECAPKSPQASIKWLQDK 609
OY 525 A-PNSRYLLSCPMESRHAITYSWR-HKEN 550

RESULT 4
ENTRY G01856 #type complete
TITLE semaphorin V - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
04-Sep-1998

```


Matches	139;	Conservative	121;	Mismatches	140;	Indels	42;	Gaps	30;
Db	171	PYDPLDTRASALINDELAVGYI--DEMGTDIAIFRTLGKQAMRTDQYNSRMUDPSFIH	229						
Oy	139	PFSPDENSLVLFEGDEVYSTIRKEIYNGK-IPFRIRIGSELYTS--DIV-MQNPFQIK	194						
Db	230	AELIPDSANDDKLTYFFERRSAAE-APQSPAVYARIGRICLINDGHCCL-VNKNSTFLK	287						
Oy	195	ATIVHQDQAYDCKIYFFREDNPDKNEAPLANSVRVAQQLCGDGGSSLSVSKMNTFLK	254						
Db	288	ARLYCSVGEDEGIEIHFDELQDVFV-QQTQDV-RNPVIYAVFTSSGSYVFGSAVCYSMA	345						
Oy	255	AMLYCS---DATATKNFNRLQIDFELLPDPSGQMDTRTYGVFSNMP-WNY--SAYCVSLG	308						
Db	346	DIRWVENGPPFAHKSGPNYQWMPFGSKMYPBPGVCPGSETFPPSKSRKMDYDDEVINFRS	405						
Oy	309	DIDKVFRTS-SLK-G--YH-----SS-LFNPRGKCLPDQ-QP-I-PIETF--QYAD--R-	351						
Db	406	HPLMQAVYPIQRRPLVYRT-GAAYRLLTIIVDDVDSADG-RYEVLFLGIDRGVQKVIY	463						
Oy	352	HPEVAQREPERNG--PLKRIPLSHSKYHQKAVHNMQASHGETFHVLYITTDRTGRIHKVYE	409						
Db	464	LPKDDQEMEELMLEVEVEKDPAPVKTWTJTSKQQLYVASAVGVTLSLHRCOAYGAC	523						
Oy	410	-PGEDEHSEFAFINIMEIOFFRRAAAIQWMSLDIAERKRKLVSWSQWESQVPLDLCIYVGGC	468						
Db	524	ADCCCLARPPYAMPGQACSRRTASKKRRSRQDVRHGNPIRQCGFGFSNMKNKNVESQY	583						
Oy	469	HGCLMSRDPYCGMQDGRGIS-IYSSEK-SYLDISINPAEPHECP--NPKDPKAPLQKVS	524						
Db	584	GVAGSAFLCQCPSPQATVKM	605						
Oy	525	A-PNSRYTLSCPMESRHAITYSM	545						
RESULT	6								
ENTRY	A49069	#type complete							
TITLE		collapsein - chicken							
ORGANISM		#formal_name Gallus gallus gallus #common_name Chicken							
DATE		07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 04-Sep-1998							
ACCESSIONS	A49069								
REFERENCE	A49069								
#authors		Lu, Y.; Raible, D.; Raper, J. A.							
#journal		Cell (1993) 75:217-227							
#title		Collapsein: a protein in brain that induces the collapse and paralysis of neuronal growth cones.							
#accession	A49069								
#status		preliminary: not compared with conceptual translation							
#molecule_type	mRNA								
#residues	1-772	#label LUO							
#cross-references	GB:U02526; NID:g410078; PID:g410079								
CLASSIFICATION		#superfamily semaphorin							
SUMMARY		#length 772 #molecular_weight 88867 #checksum 9712							
Query Match	15.5%	Score 681; DB 2; Length 772;							
Best Local Similarity	31.5%;	Pred. No. 1,09e-121;							
Matches	146;	Conservative 119; Mismatches 150; Indels 48; Gaps 38							
Db	166	RKSPDPKLLTASLIVDGLYSGTA-ADEMGDFAIFRTLGHHPIRTEOHDSRWLND	224						
Oy	135	RGYAPFSPDENSLVLFEGDEVYSTIRKEIYNGK-IPFRIRIGSELYTS--DIV-MQNP	190						
Db	225	RFSIAHLIPESDNEDBDKIYFFERENALDGHGTHKATHARIGQICKNDGFGHSL-VNKN	283						
Oy	191	QFIKATIVHQ--DQAYYDKIKIYFFREDNPDKNEAPLANSVRVAQQLCGDGGSSLSVSK	249						
Db	284	TTPLKARLIGCVPGNGIDTHFDLQDVFCL-NSKDP-KNPITYAVGTTSSNIKSGAVC	341						
Oy	250	NITFLKAMLYCS--DA-ATINKNFNLQDVFELLPDPSGQMDTRTYGVFSNPNV-Y--SAYC	303						
Db	342	MYSMIDARVRYLIGYAHNRDGPNYQWVYQGVYPRPGTGPSTKTFGGFSDTKDLPDEVIT	401						

Qy	304	VYSLDDIDKVFETS-SLK-G--YH--S--S-LPNRPCKLPDO-QPI-PRPTF--OY-A	349
Db	402	FARSHAPMNPVPIPN--RPIMKITDVDOFTQIVVDVADBDGQ-YDVMFIGTDIGTVL	459
Qy	350	-DR-HPEVAQVRERPGPLKTLPLFHSK--YHQKVAVHMQQSHGETFHVLYLTTDRGTH	405
Db	460	KVVSJPKEMHLEEVLEEMTVFERPPVISMAKISTKQOOLYIGSATGVSQLPLHRCDI	519
Qy	406	KVVE-PGDEHSHFAFNIME-IOPFRRAAIQITMSIDAEERKLKLYVSSQWEVSQVPLDCEV	463
Db	520	YGKAAECCLLADPFCAMDGSSCSYPTTAKRTRRRODIRNGDPLTTCSDLOHHNDPSGQ	579
Qy	464	YGGGCHGLMSDPLCGMDQGRCSITYSSE-RSVL-QSINPAEPKKECPN-KPRDKAP--	518
Db	580	TLEEKIIYGVENSSTFLCSPKSPQAIYV-WOFQONDDHRYE	621
Qy	519	-LQ-KVSLA-PMRRYLSC-PMESHATYMSR-KHENVEQGE	556
RESULT	7		
ENTRY		158169	#type fragment
TITLE		semaphorin III - mouse (fragment)	
ORGANISM		#formal_name Mus musculus #common_name house mouse	
DATE		26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 04-Sep-1998	
ACCESSIONS		158169	
REFERENCE		158169	
#authors		Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.; Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.	
#journal		Neuron (1995) 14:949-959	
#title		Semaphorin III can function as a selective chemorepellent to pattern sensory projections in the spinal cord.	
#cross-references		MUID:95267432	
#accession		158169	
#status		preliminary; translated from GB/EMBL/DBJ	
#molecule_type		mRNA	
#residues		1-666	#label RES
#cross-references		GB:L40484; NID:9703189; PID:9703190	
GENETICS			
#gene		SematiI	
CLASSIFICATION		superfamily semaphorin	
SUMMARY		#length 666	#checksum 9654
Query Match		15.2%;	Score 667; DB 2; Length 666;
Best Local Similarity		30.1%;	Pred. No. 1,766-118;
Matches 150;		Conservative 130;	Mismatches 168; Indels 50; Gaps 39;
Db	60	RCKSYDRKLTLASLLIGELYSGTA-ANFMGRDPAIRRTIGHHNPIRTEDHDSRWLNDP	118
Qy	135	KQYAFSPDENSLVLFEGDEYVSTRKQELNGK-IPRRRRIGSELYTS--DTV-WQNP	190
Db	119	RFISAHLIPESNPDEDDKYVFFEFRENALIDGESHGKATHRIGQICKNDGFGHRL-VNKM	177
Qy	191	QPIKATIIYHQ-DQAADKKIITYFFREDNDNDKPNPEARLANSRKAQLCRDGGESSLSIKM	249
Db	178	TFELKARLICSVPGNGIDTHFELQDYFLM-NSKDP-KNDIYVGVFTTSSNIFKGSAYC	235
Qy	250	NTEFLKAMLVYS--DA-ATNKNFNRLQDYFLPLDPDSGGQRDRYVGVSPNPN-Y--SAVC	303
Db	236	MYSMSDVARVLLGPRAHHDGPNYQWVPYQGVAPPYPRPETSCKTFGGFDSIKDLPDQYIT	295
Qy	304	VYSLDDIDKVF--F--RTSS-LK--GVHSSLPLNRPCKLPDO-QPI-PRPTF--OY-A	349
Db	296	FARSHAPMNPVPIPN--RPIMKITDVDOFTQIVVDVADBDGQ-YDVMFIGTDIGTVL	353
Qy	350	-DR-HPEVAQVRERPGPLKTLPLFHS--KYHQKVAVHMQQSHGETFHVLYLTTDRGTH	405
Db	354	KVVSJPKEMHLEEVLEEMTVFERPTTISAMELSTKQOOLYIGSTAGVQAOLPLHRCDI	413
Qy	406	KVVE-PGDEHSHFAFNIME-IOPFRRAAIQITMSIDAEERKLKLYVSSQWEVSQVPLDCEV	463
Db	414	YGKAAECCLLADPFCAMDGSSCSYPTTAKRTRRRODIRNGDPLTTCSDLOHHNDHNP	473

QY	464	YGGCGCHGLMSRDPRYCGNDGRCISITYSSE-RSVL-OSINPAEPHKECPN-PRPKD--AP	518
Db	474	SLEERIIIGVENSSTFLFECSPKSORALVYQFORNEDKEEKIMQDHIIRTEOGILLRS	533
QY	519	LQ-KVSLA-PNSRYIYLSLSC-PMERHNTY-SW-RHKENWEGSCSEPE-HQ-SPNCILFIEN	570
Db	534	LQKDSGYILCHAVEHGF	551
QY	571	LTAQGYGHYFCEACEGSGY	588
RESULT	8		
ENTRY	148747	#type complete	
TITLE	semaphorin D - mouse		
ORGANISM	#format_name Mus musculus #common_name house mouse		
DATE	02-Jul-1996 #sequence_revision 02-Jul-1996 #ext_change 04-Sep-1998		
ACCESSIONS	148747		
REFERENCE	148744		
#authors	Puschel, A.W.; Adams, R.H.; Betz, H.		
#journal	Neuron (1995) 14:941-948		
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension		
#cross-references	MU01:95267431		
#accession	148747		
#status	preliminary; translated from GB/EMBL/DBAJ		
#molecule_type	mRNA		
#residues	1-772 #label RES		
#cross-references	EMBL:X85993; RID:9854329; PID:9854330		
GENETICS			
#gene	semd		
CLASSIFICATION	#superfamily semaphorin		
SUMMARY	#length 772 #molecular_weight 88710 #checksum 1776		
Query Match	15.2%	Score 669; DB 2; length 772;	
Best Local Similarity	30.2%; Pred. No. 6,13e-119;		
Matches 170; Conservative	143; Mismatches 182; Indels 67; Gaps 51;		
Db	58	HTFLDDE-ERSRLVYGAKDHFESFNLYINIKDFOKIWPVSY-TRDECKWAGKDILKECA	115
QY	43	HTVLFHEHGSSVWVGGRKYLFDPEEGKN-AS-VRTYNIIGTKSC--LDK---RDE	95
Db	116	NFIKYLEAVYNOTFLYACGTGAFHPICTIYIEVGHHPEDNIFKLDQSHFENGSKSPYDKL	175
QY	96	NYITLLERSESG-LLACGTNARHPSCWNL-V--N--GLVPL--G--EM-RGYAFSPDE	144
Db	176	LTAASLIDGELYSTA-ADFMGNDFAFRRLGSHHRIREFQHDHSMKLNDRPRTISAHLP	234
QY	145	NSLVLFEGDEGYSTIRKOEYNGK-IPFRIRIGESLEYTS--DTV-MQNPQITIKATVHQ	200
Db	235	SDNPEDDKVVYFFFEFENAIIGESHGKATIHAIQICKNDFGGRHSL-VNKMTFLKARLIC	293
QY	201	DQAYVDKTYIFFFEDNDPKDNKPRPLVNSVSLDLCGDDGEGESLSVSKWNTFLKAMLYC	259
Db	294	SVPGPNIGDTHFDELQDVFLM-NSKDP-KNPVIVGVFTTSSNIFKGSAYCMYSMSDVRV	351
QY	260	S--DA-ATNKNFNRLLQDVFLPDPGSGQMDPRYGVGFNSPMWN-Y--SAYCVYSLGIDKX	313
Db	352	FLGPAHNDGPNYQWVYQGRVPRPGTCRSTKPFEGGFSTKDLPRDVTITFRSAPRAN	411
QY	314	FRTS-SLK-G--YR--S--S-LRPAPRGKCLPQO-QPI-PTETF--QV-A-DR-HPEVAQ	357
Db	412	PVPFIPNN-RPIMITDVNYOFTQIVVDVRYAEDGO-YDMVFISTDGVYGLKLVVSPKETW	469
QY	358	RVEPMGRPKTLPLFS--KYHYQVAVNHRQASIGELFHHYLLTLTDGSTITKKYK--PGDE	414
Db	470	HDLEEVLLKEKTYFERPTTISAMELSTQOOLYIGSTAGVADLP.LHRCDIYKKAACECL	529
QY	415	HSFANFINE-IQPRRAAAIQTMSLDIAERKTLVYSSQMEVSGVPRDLCEVYGCGCHGCLM	473
Db	530	ARDPYCAMDSSGCRVYPTAKRRTRRQDRIKNGRPLTHGSDLEDHNNHRSLEERIIIGV	589
QY	474	SRDPYCGNDGRCISITYSSE-RSVL-OSINPAEPHKECPN-PRPKD--AP-LQ-KVSLA-	525

ENTRY	9	#type complete
RESULT	D49423	semaphorin iii precursor - human
TITLE		formal name Homo sapiens #common_name man
ORGANISM		06-Jan-1993 #sequence-revision 06-Jan-1993 #text-change
DATE	04-Sep-1998	
ACCESSIONS	D49423	
REFERENCE	A49423	
#authors	Kolodkin, A.L.; Mattes, D.J.; Goodman, C.S.	
#journal	Cell (1993) 75:1389-1399	
#title	The Semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules.	
#accession	D49423	
#status	preliminary: nucleic acid sequence not shown	
#molecule_type	mRNA	
#residues	1-771 ##label KOL	
#cross-references	GB:L26081; NID:g799328; PID:g436560	
GENETICS		
#gene	GDH:SEM1	
#cross-references	GDH:283448	
CLASSIFICATION	#superfamily semaphorin	
SUMMARY	#length 771 #molecular-weight 88889 #checksum 6249	
Query Match	15.0%; Score 661; DB 2; Length 771;	
Best Local Similarity	30.3%; Pred. No. 4,156-117;	
Matches	151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;	
Db	166 RGSFYDPKLTASLLIDGELYSGA-ADYNGRDFATRTTGHNHPIRTEDHDSRWLNDP	224
Qy	135 RGYAFSPDENSELVLEGEDEVYSTIKQELYNGK-IPRRIRIGSELYTS--DIV-QQNP	190
Db	225 KFI SAHLISESDNPDDKVVYFEFFRENALDGEHSGKATARIQICLKNDFGHRSL-VNKK	283
Qy	121 QFIKATIVHQ-DQADADKIIYFFREDNDKPNRPLANSRYAQLCRGQGESLSLSKVV	249
Db	284 TTEFLKRLICSVPGNGIDTHEDELQVFLM-NFKDP-KNPVYGVFTTSSNIFKGSAYC	341
Qy	250 NTFELKAMLYCS--DA-ATNKNFNRLQDVFLLRPDSGGQRDRYRVGVFSNPMNV-Y--SAVC	303
Db	342 MYSMSDVARVFLGPRAHNDGPRYQGVNPRVPRPCTCSKFFGGDSFKDLRDPVIT	401
Qy	304 VYSLDDIDKVFETS-SLK-G-YH--S--S-LPNRPCKLPDQ-QPI-PIETP--QV-A	349
Db	402 FARSHPMYINPFPNNRPRIYKLTQVNTQFIQVVDRAVDAEDGO-YDMFGITDVTGLK	460
Qy	350 -DR-HPEVAQRREPGRPLKTRP-FLSKYHQKAVNHRQASHGELFHYLYLTTRGILHK	406
Db	461 VVSIKETWYDLEEVLEEMTVFREPRALISAMELSTKQOQLYIGSTAGVAQPLRHRCDIY	520
Qy	407 VYE-PGEDEHSFAFNIME-IQFRRRAALQIMSIDAERKRYLVSQSEVSGVPLDCEVY	464
Db	521 GKACACCLARPYCAMGWSACSRKFFPAKRTTRQDRLRNDPLRHOSDLHNDHNGHSP	580
Qy	465 GGGCGCLMSRPYCGWQGRICISLISSE-RSVL-QSINPAPRPKECPNRPDK---AP	518
Db	581 EERIIVGENSSTFLFECSPKSRQALVYQOFORNEERKEELRVVDHILIRTOGILLSRLQ	640
Qy	519 LQKVELA-PNSXYIYLSG-PMESRNTATY-SW-RHKENV-EQSCFEGHQ-SPIICLFIENLT	572
Db	641 OKDSGNYLCHAVEHGFIQ	658
Qy	573 AQQYGHYFCEAQEGSYFR	590
RESULT	10	
ENTRY	148746	#type fragment
TITLE	semaphorin C - mouse (fragment)	

```

ORGANISM      #formal_name Mus musculus #common_name house mouse
DATE          02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
              28-Feb-1997

ACCESSIONS    I48746
REFERENCE      Puschel, A.W.; Adams, R.H.; Betz, H.
#authors      Neuron (1995) 14:941-948
#journal      Murine semaphorin D/collapsin is a member of a diverse gene
#title        family and creates domains inhibitory for axonal extension
#cross-references MUID:95267431

#cross-references MUID:954327; PID:9854328
#accession    I48746
#status       preliminary: translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-782 #label RES.
#cross-references EMBL:X85992; NID:9854327; PID:9854328
GENETICS
#gene
#summary      semc
SUMMARY        #length 782 #checksum 1571

Query Match      14.0% Score 618; DB 2; Length 782;
Best Local Similarity 33.7% Pred. No. 2,66-107;
Matches 149; Conservative 96; Mismatches 148; Indels 49; Gaps 32

Db
74 KRDCCNYIKILPLPSSLLTCGTAAPSLCAVTHIASFTIADEAGVILEDKGCHPF 133
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
91 KRDCCNYIT-LLERSEDLACGTARRPSC-W-N-----LV-NGY-VPLGEKRGYAPF 140
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
134 DPNFKSTALVYDGLTYGTVS-SFGQNDPAISRSQSSRPRTKESSLNWLQDPAFVASATS 192
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
141 SPDENSLVLFESDEYVSTIRKQENYNGKIPRRRIRGESELYTSDTV--MNPQFIKATIV 198
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db
193 PESTLSPGIDDDKIYFFSETGEQEEFFENTI-VSRVARCKGGEGGERVLQ-QRMTSFL 250
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
199 HODQ-A-Y--DKIYFFREDNPD-KNEEALNLSVAQLCRGQGGGSSLSVSKWMTFL 253
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
251 KROLLCSPPDGEPFNVLQDVFETLNPQ-QMRKTLISGVTSQWHRGTTGSAICVFTM 309
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
254 KMLVCSDPATNKNFNRLQDVFLL-PDSGQKQITRYGVFSNWN--Y--SAVCVYSL 307
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db
310 NDVQKAFDGLYKKVYRETQWYETTHQVPRPGACITNSAREKKNINSLOLDPRLVNLFL 369
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
308 GDIDKVF-----RISL-LKG-Y-HSS-LPNRPQKCLPD-QQPIPETRQVADRHHVEVA 356
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db
370 KDHLMDQVNSRLLLQPRARYQVAVHYRVLGHS-TYDVLFLGTGDDGLRHKAVTLSSR 428
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
357 QVEEPM-GPLTPTLF--HSKYIKQVAAVHRMQASHGETFEHVLTYLTDRTGIHKVVEGEQ 413
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db
429 VH---I-IEEDQIFPQSGPQVNLDDSHSGGLLYXSSHGYYQVQVAVNCSLYPT-CGGCCL 483
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
414 EHSFAFNIMEIQFERRAAIQTMSLDAERRLLYVSSQVEVSQVPLDCEYVGGGCHCCLM 473
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db
484 ARDPYCAWTSACRLASLYOPD 505
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY
474 SRDPYCGMDQGRK-I-STYSSE 493
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 11
ENTRY      S66498 #type complete
TITLE      M-sema F protein precursor - mouse
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE        28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change
10-Sep-1997

ACCESSIONS S66498
REFERENCE   S66498
#authors    Inagaki, S.; Furuyama, T.; Iwahashi, Y.
#journal     FEBS Lett. (1995) 370:269-272
#title       Identification of a member of mouse semaphorin family.
#cross-references MUID:95385809
#accession   S66498
#status      preliminary
#molecule_type mRNA
#residues     1-834 #label INA
#cross-references EMBL:579463; NID:g1110598; PID:g1110599

```

FEATURE		#domain signal sequence	#status predicted	#label SIG
1-21				
22-834		#product M-sema F protein	#status predicted	#label MAT
SUMMARY		#length 834	#molecular-weight 92556	#checksum 7189
Query Match		11.9%:	Score 522;	DB 2; Length 834;
Best Local Similarity		28.5%:	Pred. No. 1,30e-85;	
Matches 142;		Conservative 118;	Mismatches 190;	Indels 49; Gaps 36;
Db	102	KGKS-NQTECFNFRFLDOPYNSHLVCGFYAQPCKTYINMLFTLDRAEFDGKGP	160	
Qy	85	KGSLCDKDCENYITLLER-RSEGLACGTNAHPSC-W-NLVNGTV--VPLEMGRIAP	139	
Db	161	YDPAGKTGTLVDLVEYSATLN-NFELGETEVLIRYMCTHHSIKETEXLAFLWNEPHEVGA	219	
Qy	140	FSPDENSLVLEPGDEGVSTIKRQYNGKIDPRFRIRIGESLEYTSN-TV-MQNPQIKAT	196	
Db	220	FVPEVSFGTFDDDKIYFFFSERAVEYDCISEOVAVARVAVCMDMGARTLD-KKWITF	278	
Qy	197	IYHD-QAY--DDKIYFFREDPDNDNPPALNVSRVALDLGGDGGESSLVSXMTF	252	
Db	279	LKARLYCASPMPKYVFNOCLKVHTLRGAS-WHNTHFEVGFAQRMGDMDLSACEYOLEO	336	
Qy	253	LKAMLYCDAATNNFNRLDVFLPPRSQMRTRYGYGSNW--NYSAYCVSYSLD	309	
Db	337	IOOVEGFKEYSEQADKMARVTDPVPSPRSGCINNMRDNQYTSSLELPNTLFIRK	396	
Qy	310	IDKVF---RT-S-SLTK-G-YHSSLPNRPGRCLPD-QQPIP-TETFOVADRHE-VA-	356	
Db	397	HPLMDQVKRPLRGRLVKKTNTHVADVRELDGATITVYFTIGTGGMILKAVSLCP	456	
Qy	357	QR-VF-PKGR-LKTPRLFHRSKY-HQOKAAVRMAQSHEGTFHVLYLTTRDGIHKVPEGE	412	
Db	457	WIH--M-VEELOVFEDDP-VESTLYSQSKVLFAGSRQLVOLSLADCTKRF-CVDCV	510	
Qy	413	QEHSAFNIMTQFPRRRAAIQTMSLAERKKLYSSQMEVSVPLDCEVYGCGCHGL	472	
Db	511	LARDPYCAMNVNTRCAVATTSGRSGSFLVOHVANLDTSKMCNOGYGIKRVISIPKNITVS	570	
Qy	473	MSRDYPQCMD-Q-QRCISIVSSER-SVL-QSINPAEPHKCCPRPKRDKA-PLQG-VSLAP	526	
Db	571	GTDLVLPCHLSSNLAAHAHW	589	
Qy	527	NSRYLLSCPMESSRHATYSW	545	
RESULT	12			
ENTRY	148745	#type complete		
TITLE	semaphorin B - mouse			
ORGANISM	#formal_name Mus musculus #common_name house mouse			
DATE	02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 28-Feb-1997			
ACCSSIONS	148745			
REFERENCE	148744			
#authors	Puschel, A.W.; Adams, R.H.; Betz, H.			
#journal	Neuron (1995) 14:941-948			
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension			
#cross-references	MUID:95267431			
#accession	148745			
#status	preliminary; translated from GB/EMBL/DBJ			
#molecule_type	mRNA			
##residues	1-760 ##label RES			
##cross-references	EMBL:X85931; NID:g854325; PID:g854326			
GENETICS				
#gene				
SUMMARY				
	semb			
Query Match		10.9%:	Score 480;	DB 2; Length 760;
Best Local Similarity		29.2%:	Pred. NO. 3.05e-76;	
Matches 157;		Conservative 121;	Mismatches 203;	Indels 57; Gaps 41;
Db	123	QCQFNFRIVSYNATHYACGTFAFSPTAFIELODSLPLRIDKVMKGCSPLTFT	182	

QY	93	DCENITTLER-RSEGLLACSTNARHBC-W-NLVNQTVP-LGE--M--RGVAPSP-D	143
Db	183	STQAVLVG-MLYSGT-MNNLGSSEPLMRITGSHPLKIDFLRWLH-ADASEVAIPS	239
QY	144	ENSLVLEFGDEYVSTIRKQELNGKIPRRRIRGSESELYSDTVQNNQFIKATIVHDDA	203
Db	240	-TQVVFEEETASEDFEELLYISRAQVCKNDVGBEKLQ-KKWTFLKAQLLCAQPC	297
QY	204	YDDKIYFFREDNPDKNEADPLNRSVAQLCRGCGESSLSVSKWMTFLKAMLYCSDAA	263
Db	208	-QLPNNIRHAYLLPADS-P-SVERIAYVFTSQVQVGTBSANCAASLNDIEVFGKY	354
QY	264	TNKNRNRLQDVFLLPDPGQNRDTRVGVESFNPNY-----SACVSLQDIDVFR-T-	316
Db	355	KELNKETSRMTTYRGSEVSPPGSCSMSPSSDKALTE-MKD-HFLMDEHYGT-PL--L	408
QY	317	SSL-K-G-----YHSSLPNRPKGCPLDQOIPIPETQVADRHREVAQRREPMPKLTP	365
Db	409	VKSGVEYTRFLAVESARGLDSSHHVMYLGISTGFLHKAVP--ODSS-AYLVEIQLSPD	465
QY	370	FHSKHYYQKVAVHMQASHGETFHVLYLTDRGTIHKHYVEGEDEHFAFNIMEIQPFR	429
Db	466	SEPVNITLQAAQAVFEGFSGGIMRVRANCYSYES-CVDCVLARPHCAMDESRCS	524
QY	430	AAAIOTMLDERRKRLIYSSQWESQVPLDLCEYGGCGHCLMSRPDYGM-DGR-CI	487
Db	525	LLSGSTPKMKODMERGENDEWCTRGPARSPRQSPOLIKEVLTVPNSILELRCPHSA	584
QY	488	STYSSERSVLQSNINPAEPKRC-P-NP--K-PDK-AP--LQKVSIA-DNSRYILSCPMSR	539
Db	565	LASYHMSGRKIKISASTIYNGSL-LTLPODYGGLY-Q--CATANGSIPVSVW	638
QY	540	HATYSWRH-KENVEQSCPEGHQSPNCILFIENLTAQOYGHYFCEAQGSY-FREAOHW	595
RESULT	13		
ENTRY	E42521	#type complete	
TITLE	A39r protein - vaccinia virus (strain Copenhagen)		
ORGANISM	Organism		
#note	host Homo sapiens (man)		
DATE	09-Nov-1990	#sequence_revision	09-Nov-1990
	08-Apr-1994	#text_change	
REFERENCES	E42521		
#authors	Johnson, G.P.		
#submission	submitted to Genbank, June 1990		
#accession	E42521		
#status	preliminary		
#molecule_type	DNA		
#residues	1-403	#label JOH	
SUMMARY	#length 403	#molecular_weight 45741	#checksum 8167
Query Match	10.7%	Score 470:	DB 2:
Best Local Similarity	33.5%	Pred. No. 5,05e-74:	Length 403:
Matches	86:	Conservative	59:
		Mismatches	95:
		Indels	17:
		Gaps	
Db	77	LVCCTGNGNCPKCMK-IDGSDDPKHRGCIYAYQNSKVTIISYN-ECVLSIDINISK-EG-I	1322
QY	109	LACGINAHHPCCMNLVNGTVPLGEMGVIAPFSPDENSLVLEFDEDEYVSTIRKQENGI	168
Db	133	KRMRFDPGCGVLYLTADNVYPKKG-IRGAFAVDKGYD-KVYTLFDTIGSKR-I-VK	187
QY	169	PRFRIRKE-S-ELTSDTVQNNQFIKATIVHDDQAYDKIYFFREDNPDKNPEAPLN	226
Db	188	IPYIAOMLBNEDGPPSSLSHRMSTFLKVELEC-DID-GRSY-R-Q-IHSRTIKTD-ND	241
QY	227	VSRAQALCRGQGGESSLSVSKWMTFLKAMLYCSDATNNFNPNLQDVFLLPDPGQWRD	266
Db	242	TILVYFEDSPYSKALCTYSKNTIKQSESTYSKLEGYTKQLPSPAPGICLPAGKVSHTTF	301
QY	287	TRYVGVEFNPNPNYAVCYISGLDIDKVFRTSSLGKGYHSSLPNRPKGCPLDQOIPIPET	346

Db 302 EVIEKYNVLDITIKPLS 318
: : : : :
QY 347 QVADRHPEVAQREVPWG 363

RESULT 14
ENTRY S29921 #type complete
TITLE hypothetical protein 15 - vaccinia virus
ORGANISM #formal_name vaccinia virus
DATE 20-Sep-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997

ACCESSIONS S29921
REFERENCE S29907
#authors Amesgadzic, B.Y.
#submission submitted to the EMBL Data Library, January 1991
#accession S29921
#status preliminary
#molecule_type DNA
#residues 1-441 #label AME
#cross-references EMBL:X57318; NID:962239; PID:962254
SUMMARY #length 441 #molecular_weight 50185 #checksum 6034

Query Match 10.5%; Score 463; DB 2; Length 441;
Best Local Similarity 33.9%; Pred. No. 1,79e-72;
Matches 87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;

Db 115 LVCGTNNGNPKCK-IGSDDPKRGHGRYAPYONSKVTIISHNGC-VLSDINISK-EG-I 170
: : : : :
QY 109 LACGTNAHNSCHNLYNGTVPLGEMRGYAPFSPDENSLVLFEGDEYVSTIRKQENYGI 168
: : : : :
Db 171 KWRREFDPCGYDLYTADNVIIPKDG-LRGAFVDKDGYD-KVYILFTDITGSKR-I-VK 225
: : : : :
QY 169 PRFRIRIGE-S-ELYSIDVMQNPQFIKATIVHODQAYDOKIYFFREDNPDKPEAPLN 226
: : : : :
Db 226 IPIAOMCLNDEGPPSSLHRMSTFLKVELC-DID-GASY-R-Q-IISRIKTD-ND 279
: : : : :
QY 227 VSRVADLCRDQGESLSVSKNNTFLKALVCSDAATNPNRLODFLLPSPSGQWRD 286
: : : : :
Db 280 TILYFEDSPYSKALCYSMNTIKOSFSTSKLEGYTKOLPSPASGICLPAGKVVPTTF 339
: : : : :
QY 287 TRYGYGFSNPNWNSAVCVSLGDIKVFRTSLKGYHSSLPNRPAGKCLPQQPIPETFT 346
: : : : :
Db 340 EVIEKYNVLDITIKPLS 356
: : : : :
QY 347 QVADRHPEVAQREVPWG 363

RESULT 15
ENTRY JH0798 #type complete
TITLE fasciclin IV precursor - American bird grasshopper
ORGANISM #formal_name Schistocerca americana #common_name American bird grasshopper
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Sep-1997

ACCESSIONS JH0798
REFERENCE JH0798
#authors Kolodkin, A.; Mathes, D.J.; O'Connor, T.P.; Patel, N.H.; Admon, A.; Bentley, D.; Goodman, C.S.
#journal Neuron (1992) 9:831-845
#title Fasciclin IV: Sequence, expression, and function during growth cone guidance in the grasshopper embryo.

#accession JH0798
#molecule_type mRNA
#residues 1-730 #label KOL
#cross-references GB:L00709; NID:g160844; PID:g160845
#experimental_source embryo
#experimental_source embryo
COMMENT This protein plays a role in growth cone guidance in the developing central nervous system.
KEYWORDS glycoprotein; transmembrane protein
FEATURE
1-22 #domain signal sequence #status predicted #label SIG\
23-730 #product fasciclin IV #status predicted #label MAT\
23-627 #domain extracellular #status predicted #label EXT\
23-627

628-652 #domain transmembrane #status predicted #label TM\
653-730 #domain intracellular #status predicted #label INT\
44,71,163,267,360,
539 #binding site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 730 #molecular_weight 81214 #checksum 5881

Query Match 10.5%; Score 461; DB 2; Length 730;
Best Local Similarity 28.0%; Pred. No. 4,96e-72;
Matches 143; Conservative 130; Mismatches 182; Indels 56; Gaps 46;

Db 32 QFGEERYQFLNESHKDHFKLEKDHNSLLGARNIVINISRLDTFTQRIENHSSG 91
: : : : :
QY 29 HVGQDRVD-FGQTEPHTVLFH--EPGSSSVWVGGRKRVLFDFPECKN-ASVPTV-NI-G 82
: : : : :
Db 92 AHRELCLYKGSSEDDCONYTRVLAKIDDDVLLCGTNAVYPLCRHVALKQGVVKEYE 151
: : : : :
QY 83 STKGSC-L-DKR--DCENYITLERSSEG-LIACGTNAHNSCHN--LVNGT-VVPLG-E 133
: : : : :
Db 152 GRGLCEPDHPDHNSTAYISEGLYSATV-ADFGSTDP-LI-YRG-P-LRTERSDL-K-QLN 204
: : : : :
QY 134 MRGYAPFSPDENSLVLFEGDEYVSTIRKQENYKIRFRIRIRESELYSDTVMQNPQFI 193
: : : : :
Db 205 APNFVN-TMEYNDIFFEFRETAVEYINCQAIY-SRVARVCKHDKGPHQFG-DRWTSF 261
: : : : :
QY 194 KATIVHODQAYDOKIYFFREDNPD-KNPEAPLVNRSVAQLCGDQGESLSVSKWNTF 252
: : : : :
Db 262 LKSLRNCVSPGDYFYFNEIQSTSDIEGNGQ-VEKILYGFYTPVNSIGSAYCAFS 320
: : : : :
QY 253 LKALVCSDAATNK-NFNRLQDFV-LLPDP-SGQMDTRVYGVSPNWN-Y--SAVCVYS 306
: : : : :
Db 321 MKSILSFDPFKEQETMNSMNLAVPSLKVPEPQCVNDSRTLDPVSVNFVSKHTMD 380
: : : : :
QY 307 L-G-D-IDKVFRT-SSLKG-YHS--SL--PNRPQKCLPQQPIPETFTQVADRHPEVA 356
: : : : :
Db 381 EAVPAEFT-RPILIRISLOYRFTKIANDQVTRTPDKAYDVLFIGTDGKVIKALNSASF 439
: : : : :
QY 357 QVEPMPGLKTPLFH-S-KVHYQKAV-HRMQASHGETFHVLYLTDRGTHIRKVEPG-- 411
: : : : :
Db 440 DSSDITVSVYIELOVLPQGVPAKNIYVYRMDGDDSKLVVSDDELLAKLHMGSGKIT 499
: : : : :
QY 412 EOHSSFAFNIME-IQFFRRALAIQT--SLDERRKLVSQWEVSOVPLDCEVYG-G 466
: : : : :
Db 500 NCRECVSLDPPYCAMDNVELCTAGSPDMS 530
: : : : :
QY 467 GCHGLMSRDPICGMDQ-G-KCISTISSEKS 495

Search completed: Thu Jul 8 19:16:58 1999
Job time : 40 secs.

THIS PAGE BLANK (USPTO)

W E S E H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch.p protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 19:13:59 1999; Maspar time 18.36 Seconds
Tabular output not generated. 906.652 Million cell updates/sec

Title: >US-09-041-236-2
Description: (18-606) from US09041236.pep (32 of 45)
Perfect Score: 4401
Sequence: 1 SGPRFAVMKGVGHGQDRVDF.....SYREQNHQMLPEDGIMAE 589

Scoring table:
PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 50.252; Variance 76.002; scale 0.661

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	470	10.7	403	1	VA39_VACCC	3.43e-86
2	463	10.5	441	1	VA39_VACCV	2.17e-84
3	163	3.7	1871	1	SEX_HUMAN	9.46e-13
4	105	2.4	227	1	US08_HCMVA	3.85e-02
5	102	2.3	275	1	NRT2_RAT	1.13e-01
6	95	2.2	255	1	HA21_HUMAN	1.24e+00
7	99	2.2	275	1	NRT1_RAT	3.21e-01
8	99	2.2	460	1	EF1A_TIRRE	3.21e-01
9	95	2.2	518	1	VL2_HPV5B	1.24e+00
10	95	2.2	518	1	VL2_HPV5B	1.24e+00
11	99	2.2	591	1	VRP2_SALTY	3.21e-01
12	99	2.2	591	1	VRP2_SALTY	3.21e-01
13	99	2.2	591	1	VRP2_SALTY	3.21e-01
14	99	2.2	593	1	VRP2_SALTY	3.21e-01
15	95	2.2	683	1	AMOH_ARTGO	1.24e+00
16	99	2.2	916	1	PMAL_AJECA	4.52e-01
17	98	2.2	1132	1	DNE1_HSV6U	3.21e-01
18	95	2.2	1390	1	MET_HUMAN	1.72e+00
19	94	2.1	329	1	RBSF_ECOLI	4.36e+00
20	91	2.1	348	1	DDL_ENTFA	4.36e+00
21	91	2.1	356	1	VP3_NPVLD	4.36e+00
22	93	2.1	380	1	DP3B_MYCPN	2.39e+00
23	92	2.1	442	1	CEM1_YEAST	3.31e+00

24	92	2.1	460	1	EF1A_NEDCR	3.31e+00
25	91	2.1	498	1	ELAS_PSEAE	4.56e+00
26	91	2.1	508	1	NOT_HUMAN	4.56e+00
27	91	2.1	797	1	GCYR_HCMVA	4.56e+00
28	93	2.1	827	1	Y223_METUA	2.39e+00
29	94	2.1	947	1	PM22_YEAST	1.73e+00
30	94	2.1	959	1	N100_YEAST	1.73e+00
31	94	2.1	982	1	POL_HTLV2	1.73e+00
32	93	2.1	1029	1	END1_YEAST	2.39e+00
33	94	2.1	1069	1	ENTK_MOUSE	1.73e+00
34	94	2.1	1115	1	IRE1_YEAST	1.73e+00
35	92	2.1	1132	1	DNE1_HSV6U	3.31e+00
36	94	2.1	1200	1	DDX8_CAEEL	3.31e+00
37	92	2.1	1302	1	MDR5_DROME	3.31e+00
38	92	2.1	1663	1	CO3_MOUSE	4.56e+00
39	91	2.1	1666	1	CO3_MOUSE	4.56e+00
40	90	2.0	119	1	RL19_MYCPN	6.25e+00
41	90	2.0	213	1	AMEX_BOVIN	6.25e+00
42	90	2.0	254	1	HA22_HUMAN	6.25e+00
43	89	2.0	649	1	RA32_SCHPO	8.55e+00
44	89	2.0	966	1	AMEN_HUMAN	8.55e+00
45	90	2.0	1874	1	POLR_XYMYV	6.25e+00

ALIGNMENTS

RESULT	1	STANDARD	PRT	403 AA
ID	VA39_VACCC			
AC	P21062			
DT	01-FEB-1991 (REL. 17, CREATED)			
DT	01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)			
DE	01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)			
DE	PROTEIN A39.			
GN	A39R.			
OS	VACCINIA VIRUS (STRAIN COPENHAGEN).			
OC	VIRUSES: DSDNA VIRUSES, NO RNA STAGE: POXVIRIDAE; CHORODOPOXVIRINAE;			
OC	ORTHOPOXVIRUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 91021027.			
RA	GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,			
RA	PAOLETTI E.;			
RT	"The complete DNA sequence of vaccinia virus.";			
RL	VIROLOGY 179:247-266(1990).			
RN	[2]			
RP	COMPLETE GENOME.			
RA	GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,			
RA	PAOLETTI E.;			
RL	VIROLOGY 179:517-563(1990).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: M35027; G335517; -			
DR	PIR: E42521; E42521; -			
SQ	SEQUENCE 403 AA; 45741 MW; FE0DB9AB CRC32;			
Query Match	10.7%; Score 470; DB 1; Length 403;			
Best Local Similarity	33.5%; Pred. No. 3.43e-86;			
Matches	86; Conservative 59; Mismatches 95; Indels 17; Gaps 16;			
DB	77 LVCGTNGNGPKCKW-IDSDDPKRGKRGAPQNSKVIIISYN-BCVSDINISK-EG-I 132			
OY	109 LACGTHARHPSCWNLVNGVPLGMRGAPSPSPNSVLVEGDEVSTIKQKQYNGIT 168			
DB	133 KWRMRFDSPGCDLYTADNVIPKDG-LGAFYDKDGYD-KVYILFTPLIGSKR--I-VK 187			
OY	169 PFRFRIRGE-S-ELVTSVTQMNPOFIKATIVHQQAADKIKYFRFRDNPKNEARPLN 226			

Db 188 IPIYAQMCLNDEGGPSSLSHRWSTFLKVELEC-DID-GRSY-R-Q-IIHSRTIKTD-ND 241
Y 227 VSRVAOLCRGDOGGESSLSVSKMNTFLKAMLVCSDAATNKNENRLODVFLLPDPGQWRD 286
Db 242 TLIVFEFSPYSKSLCTYSNMNTIKOSFSTSKLEGYTKOLPSPAGTCLPAGKVVSHTF 301
Y 287 TRVYGVFSNPNWNSAVCYVSLGIDIKVFTSRLKGYHSSLPNRPBGKCLPDPQPIPTETF 346
Db 302 EVIERKYNVLDDIIRPLS 318
Y 347 QVADRHPVAGORVEPMG 363
RESULT 2 STANDARD: PRT: 441 AA.
AC P24764:
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE PROTEIN A39.
GN A39R OR (SALL9R AND SALL9R).
OS VACCINIA VIRUS (STRAIN WR).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORODOPOXVIRINAE;
OC ORTHOPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91310644.
RA AMEGADZIE B.Y., AHN B.-Y., MOSS B.;
RT "Identification, sequence, and expression of the gene encoding a Mr
35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase.";
RL J. BIOL. CHEM. 266:13712-13718(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91259063.
RA SMITH G.L., CHAN Y.S., HOWARD S.T.;
RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
the right inverted terminal repeat.";
RL J. GEN. VIROL. 72:1349-1376(1991).
CC -1- CAUTION: IN REF.2 THIS SEQUENCE IS IN TWO PART (SALL9R AND
SALL9R) DUE TO THE PRESENCE OF FRAMESHIFT COMPARED TO THAT OF
REF.1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M61187; G335798; -;
DR EMBL: D11079; G322726; ALT_SEQ.
DR EMBL: X57318; G62254; -;
DR PTR: S29921; S29921.
SQ SEQUENCE 441 AA; 50185 MW; 2C82A68 CRC32;
Query Match 10.5%; Score 463; DB 1; Length 441;
Best Local Similarity 33.9%; Pred. No. 2,176-84;
Matches 87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;

Db 280 TLIVFEFDSYKSLCTYSNMNTIKOSFSTSKLEGYTKOLPSPAGTCLPAGKVVPHTF 339
Y 287 TRVYGVFSNPNWNSAVCYVSLGIDIKVFTSRLKGYHSSLPNRPBGKCLPDPQPIPTETF 346
Db 340 EVIERKYNVLDDIIRPLS 356
Y 347 QVADRHPVAGORVEPMG 363
RESULT 3 STANDARD: PRT: 1871 AA.
ID SEX_HUMAN
AC P51805;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE TRANSMEMBRANE PROTEIN SEX PRECURSOR.
GN SEX.
OS HOMO SAPIENS (HUMAN).
OC EUMAROTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-FETAL BRAIN, SKELETAL MUSCLE, AND EMBRYO;
RX MEDLINE: 96149362.
RA MAESTRINI E., TAMAGNONE L., LONGATI P., CREMONA O., GUISANO M.,
RA BIONE S., TAMANINI F., NEEL B.G., TONIOLO D., COMOGGIO P.M.;
RT "A family of transmembrane proteins with homology to the
MET-hepatocyte growth factor receptor";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:674-678(1996).
CC -1- FUNCTION: POTATIVE RECEPTOR INVOLVED IN THE DEVELOPMENT OF NEURAL
CC AND EPITHELIAL TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -1- TISSUE SPECIFICITY: PREDOMINANT IN NEURAL TISSUE DURING
CC DEVELOPMENT (BY SIMILARITY). WIDELY EXPRESSED IN ADULT TISSUE.
CC -1- SIMILARITY: WITH THE EXTRACELLULAR DOMAIN OF THE MET/ROH/AGF
CC RECEPTORS AND THE CYTOPLASMIC DOMAIN OF SEP, NOV AND OCT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X87852; E183847; -;
DR MIM: 300022; -;
DR RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
KW SIGNAL
FT 1 19
FT CHAIN 20 1871
FT DOMAIN 20 1871
FT TRANSMEM 1221 1241
FT TRANSMEM 1242 1871
FT DOMAIN 39 59
FT CARBOHYD 548 548
FT CARBOHYD 637 637
FT CARBOHYD 738 738
FT CARBOHYD 746 746
FT CARBOHYD 1009 1009
FT CARBOHYD 1036 1036
FT CARBOHYD 1073 1073
FT CARBOHYD 1115 1115
FT CARBOHYD 1162 1162
SQ SEQUENCE 1871 AA; 207661 MW; 9A11046A CRC32;
Query Match 3.7%; Score 163; DB 1; Length 1871;
Best Local Similarity 29.5%; Pred. No. 9,46e-13;
Matches 26; Conservative 29; Mismatches 29; Indels 4; Gaps 3;


```

D6      484 SOLPVTCEQKOS--CAACTGSGDCPGM 510
          |||:| | :| | | | | | | | |
OY      454 SQVPLDLCEVIGGGCHGLMSRDPICGW 481

RESULT      4              STANDARD:      PRT:      227 AA.
ID US08 HCMVA
AC P09730.
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN HXLFF4 PRECURSOR.
GN US8.
CS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OS VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87169717.
RA WESTON K., BARRELL B.G.;
RT "Sequence of the short unique region, short repeats, and part of the
RT long repeats of human cytomegalovirus."
RL J. MOL. BIOL. 192:177-208(1986).
[2]
RP COMPLETE GENOME.
RX MEDLINE: 90269039.
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORSNELL T., HUTCHISON C.A. I II, KOUDARIDES T., MARTIGNETTI J.A.,
RA PRIDDLE E., SATCHEL W.S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL CURR. TOP. MICROBIOL. IMMUNOL. 194:125-169(1990).
- - SIMILARITY: BELONGS TO US6 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost on
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X17403; G59767; -.
DR EMBL: X04650; E3282; ALT_SEQ.
DR PIR: 126078; Q0BEC9.
DR PIR: S09922; S09922.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE; SIGNAL.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 227 HYPOTHETICAL PROTEIN HXLFF4.
FT TRANSEM 179 199 POTENTIAL.
FT CARBOHYD 61 61 POTENTIAL.
SQ SEQUENCE 227 AA; 26633 MW; 4662A415 CRC32;

Query Match 2.4%; Score 105; DB 1; Length 227;
Best Local Similarity 39.3%; Pid. No. 3,85e-02;
Matches 11; Conservativity 8; Mismatches 9; Indels 0; Gaps 0;

Db 17 TLAHNGNPYEDDDYYRYREDPRONGEP 44
   | | :| | | | | :| |
OY 196 TIWHDQAIVDKIYTFREDNDPKNPEA 223

RESULT      5              STANDARD:      PRT:      275 AA.
ID NRT2 RAT
AC P20974; P97912; Q95576.
DT 01-FEB-1991 (REL. 17, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE T-CELL ECOT-ADP-RIBOSYLTRANSFERASE 2 PRECURSOR (EC 2.4.2.31) (T-CELL
DE MAD(P)(+)-ARGININE ADP-RIBOSYLTRANSFERASE 2) (T-CELL MONO(ADP-
DE RIBOSYL)TRANSFERASE 2) (ALLONANTIGEN RT6.2) (T-CELL SURFACE PROTEIN
DE RT6.2).
GN ART2B OR RT6-B.
```

```

OS      RATTUS NORVEGICUS (RAT) .
OC      EUCARYOTA; METAEOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC      RODENTIA; SCIUROGNATHI; MORIDAE; MURINAE; RATTUS.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 90138997.
RA      KOCH F., HAAG F., KASHAN A., THIELE H.-G.;
RT      "Primary structure of rat Rt6.2, a nonglycosylated
RL      phosphatidylinositol-linked surface marker of postthymic T cells.";
RN      [2]
RP      SEQUENCE OF 1-201 FROM N.A.
RX      STRAIN-DA; TISSUE-SPLEEN;
RA      MEDLINE; 96355006.
RA      HAAG F., KUHLENAUMER G., KOCH-NOLTE F., WINGENDER E., THIELE H.-G.;
RT      "Structure of the gene encoding the rat T cell ecto-ADP-
RL      ribosyltransferase Rt6.";
RN      [3]
RP      IMMUNOL. 157:2022-2030(1996).
RN      [4]
RP      CHARACTERIZATION.
RX      MEDLINE; 94193612.
RA      TAKADA T., IIDA K., MOSS J.;
RT      "Expression of NAD glycohydrolase activity by rat mammary
RL      adenocarcinoma cells transformed with rat T cell alloantigen Rt6.2.";
RN      J. BIOL. CHEM. 269:9420-9423(1994).
CC      -1- FUNCTION: HAS BOTH NAD+ GLYCOHYDROLASE AND ADP-RIBOSYLTRANSFERASE
CC      ACTIVITY (TO A LESSER EXTENT).
CC      -1- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
CC      N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC      -1- CATALYTIC ACTIVITY: NAD(+) + H(2)O = NICOTINAMIDE + ADP-RIBOSE.
CC      -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC      -1- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC      -1- PFM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC      -1- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC      FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M85193; G306806; -
DR      EMBL; X99123; E257753; -
DR      EMBL; X99122; E257751; -
DR      PIR; A34866; A34866.
DR      PROSITE; PS01291; ART; 1.
DR      PFM; PFO1129; ART; 1.
KW      TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
KW      T-CELL DIFFERENTIATION; GPI-ANCHOR.
FT      SIGNAL
FT      CHAIN
FT      PROPEP
FT      LIPID
FT      ACT_SITE
FT      SEQUENCE
SQ      246 246 246
      209 209
      31438 MW; DBEAB84E CRC32;
Query Match 2.3%; Score 102; DB 1; Length 275;
Best local Similarity 30.6%; Pred No. 1,13e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;

```

QY 131 LGE-MGCIAPFSPDENSIVLFEGDEVYSTIRKOEYNGK-IPRRIRIGESE-LYTSPTVM 187
Db 252 ES 253
QY 188 ON 189

RESULT 6
ID HA21_HUMAN STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DE 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA:
OC PRIMATES: CATARRHINI: HOMINIDAE: HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84168117.
RA AUFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPES D.,
STROMINGER J.L.;
RT "Isotypic and allotypic variation of human class II
histocompatibility antigen alpha-chain genes.";
RL NATURE 308:327-333(1984).
RN [2]
RP SEQUENCE OF 40-255 FROM N.A.
RX MEDLINE: 83065171.
RA AUFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT "cDNA clone for the heavy chain of the human B cell alloantigen DC1:
strong sequence homology to the HLA-DR heavy chain.";
RL PROC. NATL. ACADE. SCI. U.S.A. 79:6337-6341(1982).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC
DR EMBL: J00199; -, NOT_ANNOTATED_CDS.
DR PIR: A02212; HLHUC.
DR PROSITE: PS00290; IG_MHC. 1.
DR PFAM: PF00047; Ig_1.
DR PFAM: PF00993; MHC_II_alpha. 1.
DR HSSP: P01910; IIAK.
KW MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 23
FT CHAIN 24 255
FT FT 24 255 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT FT 24 255 DQ(1) ALPHA CHAIN.
FT FT 111 204 EXTRACELLULAR ALPHA-1.
FT FT 205 217 EXTRACELLULAR ALPHA-2.
FT FT 218 240 CONNECTING PEPTIDE.
FT FT 241 255 CYTOPLASMIC TAIL.
FT FT 133 189 BY SIMILARITY.
FT FT 104 104 POTENTIAL.
FT FT 144 144 POTENTIAL.
SQ SEQUENCE 255 AA; 28105 MW; 5C46DD59 CRC32;

Query Match 2.2%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1,24e+00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;

Db 25 DIVADSVAGLG-VNLQSYGSPSGQYSHEDGDGEFYVDLERKKTWQLPLFRFR 78
QY 122 NLVNGTVVPLGEMRGYAFSPDENSIVLFEGDEV-YSTIRKOEYNGKIPRRIR 175

RESULT 7
ID NRT1_RAT STANDARD; PRT; 275 AA.

AC P17982;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 PRECURSOR (EC 2.4.2.31) (T-CELL
DE NAD(P)+)--ARGININE ADP-RIBOSYLTRANSFERASE 1 (T-CELL MONO(ADP-
DE RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN
DE RT6.1).
GN ART2A OR RT6-A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA:
OC RODENTIA: SCIROGATHI: MURIDAE: MURINAE: RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEWIS A;
RX MEDLINE: 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell
RT alloantigen RT6.1.";
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
RN [2]
RP MUTAGENESIS OF GLN-207.
RX MEDLINE: 96275529.
RA MAERAMA T., HOSHINO S.-I., KATADA T.;
RT "Increase in ADP-ribosyltransferase activity of rat T lymphocyte
RT alloantigen RT6.1 by a single amino acid mutation.";
RL FEBS LETT. 388:189-191(1996).
CC -1- FUNCTION: HAS NAD+ GLYCOPOLYMERASE ACTIVITY AND EXTREMELY LOW ADP-
CC RIBOSYLTRANSFERASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
CC N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -1- CATALYTIC ACTIVITY: NAD(+) + H(2)O = NICOTINAMIDE + ADP-RIBOSE.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -1- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC
DR EMBL: X52082; G57168; -
DR EMBL: M31138; G206804; -
DR PIR: S08464; S08464.
DR PROSITE: PS01291; ART; 1.
DR PFAM: PF01129; ART; 1.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
FT T-CELL DIFFERENTIATION; GPI-ANCHOR.
FT SIGNAL 1 20
FT CHAIN 21 246
FT FT 21 246 T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
FT FT 247 275 HYDROPHOBIC, REMOVED DURING MATURATION
FT FT PROPEP 247 275 (BY SIMILARITY).
FT FT LIPID 246 246 GPI-ANCHOR (BY SIMILARITY).
FT FT ACT_SITE 209 209 BY SIMILARITY.
FT FT CARBOHYD 58 58 POTENTIAL.
FT FT MUTAGEN 207 207 O->E: INCREASED ADP-RIBOSYLTRANSFERASE
FT FT ACTIVITY.
SQ SEQUENCE 275 AA; 31388 MW; 523BLA84 CRC32;

Query Match 2.2%; Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 3,21e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;

Db 194 LGVYIKFES-FPPDDEE-LVINGEYVQKVRQNGNEIFLOSPKRRKSNYNCLSSACT 251
QY 131 LGE-MGCIAPFSPDENSIVLFEGDEVYSTIRKOEYNGK-IPRRIRIGESE-LYTSPTVM 187
Db 252 ES 253

QY 188 QN 189

RESULT 8
ID EPIA-TRIRE STANDARD: PRT: 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TER1.
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOCYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-QM9414 / RUTC-30;
RX MEDLINE: 94124021.
RA NAKARI T., ALATVALO E., PENTTILA M.:
RT Isolation of Trichoderma reesei genes highly expressed on glucose-
RT containing media: characterization of the tel1 gene encoding
RT translation elongation factor 1 alpha.";
RL GENE 136:313-318(1993).
CC - FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC - SUBCELLULAR LOCATION: CYTOPLASMIC.
CC - SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z33012; G312887; -.
DR PIR: S35772; S35772.
DR PROSITE: P500301; EFACITOR_GTP; 1.
DR PFAM: PF00009; GTP_EFTU; 1.
DR HSSP: P07157; IAIIP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 49830 MW; B9AHABR2 CRC32;
Query Match
Best Local Similarity 3.3%; Score 99; DB 1; Length 460;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;
Db 210 YGWEKETNAGKFTG-KTLLLEADISIEPKR-PTDRPLRLPLDDV 252
OY 478 YCGMDGRCISITYSERSVLOSINPAEPKKECNPRKAPLQKV 522

RESULT 9
ID VL2-HPV5B STANDARD: PRT: 518 AA.
AC P26540;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MINOR CAPSID PROTEIN L2.
GN L2.
OS HUMAN PAPILLOMAVIRUS TYPE 5B.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91306467.
YABE Y., SAKAI A., HITSUMOTO T., KATO H., OGURA H.;

RT "A subtype of human papillomavirus 5 (HPV-5b) and its subgenomic
RT segment amplified in a carcinoma: nucleotide sequences and genomic
RT organizations.";
RL VIROLOGY 183:793-798(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D90252; D1015003; -.
DR PIR: H40480; P2MLB5.
DR PFAM: PF00513; late_protein_L2; 1.
KW COAT PROTEIN; LATE PROTEIN.
SQ SEQUENCE 518 AA; 56836 MW; D0ADC243 CRC32;
Query Match
Best Local Similarity 2.7%; Score 95; DB 1; Length 518;
Matches 16; Conservative 17; Mismatches 25; Indels 1; Gaps 1;
Db 219 GDITDIELEIPSRYTEIEEPTPPRSSTPLPRNOSVGRRGFSLTNRRLVQOYVD 277
OY 402 GTIHKVEPGEQHSFAPNIMEIOPFRRAAIQTMSLDAERKKLY-VSSQWEVSQVPLD 459

RESULT 10
ID VL2-HPV05 STANDARD: PRT: 518 AA.
AC P06918;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MINOR CAPSID PROTEIN L2.
GN L2.
OS HUMAN PAPILLOMAVIRUS TYPE 5.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87207670.
RA ZACHOW K.R., OSTROW R.S., FARAS A.J.:
RT Nucleotide sequence and genome organization of human papillomavirus
RT type 5.";
RL VIROLOGY 158:251-254(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M17463; G484221; -.
DR PIR: B26277; P2ML5.
DR PFAM: PF00513; late_protein_L2; 1.
KW COAT PROTEIN; LATE PROTEIN.
SQ SEQUENCE 518 AA; 56704 MW; B7199004 CRC32;
Query Match
Best Local Similarity 2.7%; Score 95; DB 1; Length 518;
Matches 16; Conservative 17; Mismatches 25; Indels 1; Gaps 1;
Db 219 GDITDIELEIPSRYTEIEEPTPPRSSTPLPRNOSVGRRGFSLTNRRLVQOYVD 277
OY 402 GTIHKVEPGEQHSFAPNIMEIOPFRRAAIQTMSLDAERKKLY-VSSQWEVSQVPLD 459

RESULT 11
ID VRP2-SALT STANDARD: PRT: 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN MKAA.
OS SALMONELLA TYPHMURIUM.
OC PLASMID 96 KB VIRULENCE PEX102.
CC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TM: R66;
RX MEDLINE; 90136009.
RA TAIRA S., RHEN M.;
RT "Identification and genetic analysis of mkaA -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth."
RL MICROB. PATHOG. 7:165-173(1989).
RN [2]
RP SEQUENCE OF 1-10.
RX MEDLINE; 91244158.
RA TAIRA S., BAUMANN M., RIKONEN P., SUKUPOLVI S., RHEN M.;
RT "Antino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium."
RL FEBS MICROBIOL. LETT. 61:319-323(1991).
CC -1- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -1- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15042; G47783; -.
DR PIR; A54540; A54540.
KM PLASMID; VIRULENCE.
FT DOMAIN 367
SO SEQUENCE 591 AA; 65440 MW; B69F25C3 CRC32;
Query Match 2.2%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.21e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLKTKTAARLSDPOAASHTAOW-LVESYTPAGE-HIYYSILAENGNDVNL 219
QY 92 RDCENYITLLERSEGLACGTNARHPSCWNLVGTVPPLGEMGYAPSPDENSLVLE 151
DB 220 GNEAGRDRSAMRYLSKV-QYGNATPADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRIRRESELY 181
RESULT 12
ID VRP2_SALCH STANDARD; PRT; 591 AA.
AC P17450;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN (PROTEIN M2, IN MBA REGION).
OS SALMONELLA CHOLERAE-SUTIS (SALMONELLA ENTERICA).
OC PLASMID PDCSC50.
CC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RF-1;
RX MEDLINE; 90245675.
RA MATSUI H.;
RT "Nucleotide sequences of genes encoding 32 kDa and 70 kDa

RT polypeptides in mba region of the virulence plasmid, pKDC50, of
RT Salmonella choleraesuis.";
RL NUCLEIC ACIDS RES. 18:2181-2181(1990).
CC -1- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -1- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52035; G46898; -.
DR PIR; S09498; S09498.
KM PLASMID; VIRULENCE.
FT DOMAIN 367
SO SEQUENCE 591 AA; 65341 MW; BE4A439A CRC32;
Query Match 2.2%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.21e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLKTKTAARLSDPOAASHTAOW-LVESYTPAGE-HIYYSILAENGNDVNL 219
QY 92 RDCENYITLLERSEGLACGTNARHPSCWNLVGTVPPLGEMGYAPSPDENSLVLE 151
DB 220 GNEAGRDRSAMRYLSKV-QYGNATPADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRIRRESELY 181
RESULT 13
ID VRP2_SALEN STANDARD; PRT; 591 AA.
AC P55220;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN SPVB.
OS SALMONELLA ENTERITIDIS.
OC PLASMID PNL2001.
CC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ALI190;
RX MEDLINE; 94362897.
RA SUZUKI S., KOMASE K., MATSUI H., ABE A., KAWAHARA K., TAMURA Y.,
RA KIJIMA M., DANBARA H., NAKAMURA M., SATO S.;
RT "Virulence region of plasmid PNL2001 of salmonella enteritidis."
RL MICROBIOLOGY 140:1307-1318(1994).
CC -1- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -1- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14490; G517164; -.
DR PIR; S09498; S09498.
KM PLASMID; VIRULENCE.
FT DOMAIN 367
SO SEQUENCE 591 AA; 65350 MW; 9C5B2EE3 CRC32;

THIS PAGE BLANK (USPTO)

Db	350	SLKGGHSSLPNRPKGLCPDQDRIPIETQVADRHNEVQAKRVEPMKPLTLPFLHSHYHQ	409
Oy	318	SLKGGHSSLPNRPKGLCPDQDRIPIETQVADRHNEVQAKRVEPMKPLTLPFLHSHYHQ	377
Db	410	KVAVHRMQASHGETFHVLTLDRCGIHKRVBERGEQESHFAFIMELIOFFRRALAIQMS	469
Oy	378	KVAVHRMQASHGETFHVLTLDRCGIHKRVBERGEQESHFAFIMELIOFFRRALAIQMS	437
Db	470	LDAERKRLYSSQWEVSQYRPDLCLCEYVGGGCHGLMSRDPYCGMDGRCISTYSRSRYL	529
Oy	438	LDAERKRLYSSQWEVSQYRPDLCEYVGGGCHGLMSRDPYCGMDGRCISTYSRSRYL	497
Db	530	OSINPAEPHKCEPNKPDKAPLQKVSIAENSRYLYSCSPMESHRHATYSWRHKEVNEQSCP	589
Oy	498	OSINPAEPHKCEPNKPDKAPLQKVSIAENSRYLYSCSPMESHRHATYSWRHKEVNEQSCP	557
Db	550	GHQSPNCILFIENLTAAQOYIGHYFCENQDESYFREAQHNOLLPBDGIMAE	638
Oy	558	GHQSPNCILFIENLTAAQOYIGHYFCENQDESYFREAQHNOLLPBDGIMAE	606

RESULT	2		
ID	088371	PRELIMINARY:	PRT: 393 AA.
AC	088371.		
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)	
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)	
DE	SEMAPHORIN L (FRAGMENT).		
GN	SEMAPL.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUMAROTIA: MEALAZA; CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA: RODENTIA.		
OC	SCIUROGNATHI: MURIDAE: MURINAE: MOS.		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 98389619.		
RA	"LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;		
RT	"Non cytokinetic semaphorins with close homology to semaphorins of DNA		
RT	viruses".		
RL	GENOMICS 51:340-350(1998).		
DR	EMBL: AF0030699; G3523117; -.		
FT	NON_TER 393		
SO	SEQUENCE 393 AA: 43909 MM; 7CB8193C CRC32;		

Query Match	52.4%;	Score 2307;	DB 11;	Length 393;
Best Local Similarity	91.1%;	Pred. No. 0.00e+00;		
Matches	316;	Conservative	14;	Mismatches 13;
			Indels	4;
			Gaps	2

D	b	50	SQPRISAWKGG---ODHVEOSPEPTLYLFHEGGSVWVGKGKKYHHNPFEGKAASR	106
O	y	18	SGPRLFAWKGVAGDGRDEFGOTERTPLYLFEHPGSSVWVGGRKKYLDFEPGKAASR	77
D	b	107	TVINISTGSCDKDCGNYITLLERNGNLLVCYTNAKPSGCWNLVNDSVMSJGEMKG	166
O	y	78	TVINISTGSCLDKRDCENYITLLERRSGLLACGNNAHNSPCWNLNVNTGYV-PLIEEM	136
D	b	167	YAFSPDENSLVFEGDEVYSTIRKOEYNGKILPRFRIGESELYSDIVMNONPOEIKAT	226
O	y	137	YAFSPDENSLVFEGDEVYSTIRKOEYNGKILPRFRIGESELYSDIVMNONPOEIKAT	196
D	b	227	IYHQDAIDTKIYFFREDNPDKNEAPLVNSRVAOLCRGDGGESSLVSXKWNTFLKAM	286
O	y	197	IYHQDAIDDKIYFFREDNPDKNEAPLVNSRVAOLCRGDGGESSLVSXKWNTFLKAM	256
D	b	287	LVCSDAATNRNENRLQDYVLLPDPSGOMEDTINYGVFSNMNYSAACYVSLGIDIVEFT	346
O	y	257	LVCSDAATNNKNRNRLQDYVLLPDPSGOMEDTINYGVFSNMNYSAACYVSLGIDIVEFT	316
D	b	347	SSLKGIYHGCLSNPRPGMKLPKQPIETETFOYADSHPEVARQVERPMG	393
O	y	317	SSLKGIYHSSLPNPRGCKLPDQAPIETETFOYADRIPAEARREPMG	363

```

RESULT 3
AC 064906 PRELIMINARY; PRT; 653 AA.
DT 01-NOV-1996 (TREMBL:REL. 01, CREATED)
DT 01-NOV-1996 (TREMBL:REL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBL:REL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALCEPHAPINE HERPESVIRUS 1.
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RC MEDLINE: 97201573.
RA ENSER A., FLECKENSTEIN B.;
RT "Alcephapine herpesvirus type 1 has a semaphorin-like gene."
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RC MEDLINE: 97404659.
RA ENSER A., PELANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcephapine herpesvirus 1 genome."
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RA ENSER A., PELANZ R., FLECKENSTEIN B.;
RT SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U18243; G1000717; -.
DR EMBL: AF005370; G2337970; -.
SO SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

```

Query Match	42.1%;	Score 1854;	DB 14;	Length 653;
Best Local Similarity	46.98;	Pred. No. 0.00e+00;		
Matches	269;	Conservative	97;	Mismatches 195;
			Indels	12;
			Gaps	12

[illegible]

Db	608	CHATHKNDCTLLANSTTANGTHVCNMRKDS	640
QY	555	CEPGHSPNCILFLENTLAQGYGHFYCEAOGS	587
RESULT	4		
AC	062177	PRELIMINARY:	PRT: 748 AA.
AD	062177		
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)	
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)	
DE	SEMAPHORIN A PRECURSOR (SEM A).		
GN	SEMA OR SEMA.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUDAROTA; METAQOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	STRAIN-NMRT.		
RX	MEDLINE: 95267431.		
RA	PUSCHEL A.W., ADAMS R.H., BETZ H.;		
RT	"Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.";		
RL	NEURON 14:941-948(1995).		
CC	-1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.		
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).		
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.		
CC	LOW LEVELS FOUND BETWEEN DAYS 10-12.		
CC	EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL BIRTH.		
CC	-1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.		
CC	CONTAINS ONE C2-LIKE DOMAIN.		
CC	EMBL: X85990; G854324; .		
DR	MGD: MGI:107561; SEMAA.		
DR	PFAM: PF00047; 19: 1.		
KM	SIGNAL: IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;		
KM	DEVELOPMENTAL PROTEIN.		
FT	SIGNAL 1 26	POTENTIAL.	
FT	CHAIN 27 748	SEMAPHORIN A.	
FT	DOMAIN 586 649	IG-LIKE C2-TYPE DOMAIN.	
SQ	SEQUENCE 748 AA: 82894 MM; A7E53A8D CRC32:		
Query Match	16.7%;	Score 737;	DB 11; Length 748;
Best Local Similarity	33.7%;	Pred. No. 2,32e-154;	
Matches 168;	Conservative 112;	Mismatches 166;	Indels 52; Gaps 36;
Db	112	ECMNVRLAHVNHHLACRTGAHPRICALMRAATAGTASTSGPEKLEDGKKTPTDP	171
QY	93	DCENTITLLERRSEG-LIACGTNRHPSC---NLVNGT-V-V-P-LGEMRGYAFESP	142
Db	172	RHRPVSVEGLYSGV-TADMGDEIFERSLQNPSELTREPHDSRLNPKVKYFWI	230
QY	143	DENSLVLEEGEVDSTIKOEYNG-IREFRIRGESELY-S-DYV-MONPOIKAT-I	197
Db	231	PESENDDDKIYFFRESAVEAAPAMGRMSYRGQICRNDLQGRSL-VKAKTTFPKAR	289
QY	198	VHODAYVDKTIYFFRENDPNKPE-ALANVSRAVQLCRGQGGESSISVSKNMTFLKAR	256
Db	230	LVCYSPVEGDTHPDQDVFLLS-SR-DROPTLYAVFTSSGVFOGSAYCVYSMNDVR	347
QY	257	LVCS--DAATKNKNENRLDVFLLDPDSCGMQMDTRVYGVFSNPMN-Y--SAVCYSLDID	311
Db	348	RAFLGLRPHKSGPTHWYSYOGRVYPRPGMGCPKSTGTSTSTADFPDQVYQFRNPLM	407
QY	312	KVFRTS-SLK-G-----YHSSLNPPRGKCLPDQ-QPIP-TTF--QVAD--R-RPEV	355
Db	408	YNPLVPMGC-R-PLFLYAGAGYTFQIADAAADAGH-YVLEFGIDVGVLEVISVPK	464
QY	356	AQRVEPKPPLKTPLE-H--SKYHOKAVAHKMQASHGTEFVVLITTDKGTIHKVVE-P-	410
Db	465	GRNPSEGLLEELQVDFSDSAITSMQISSKROOLYVASRAVAQAIALHRCATGACACAE	524

Qy	411	GEQHSFNFNTMEIOPFRRAAIOTMSLDAERKRLYYSSOMEVSQVLDLCEVYGGGCHG	410
Db	535	CCLANDPYCAMDGSCACRFPOTAKRRFRRODIRNGDPSLCSGSSSHVLEKKVLGAVES	584
Oy	471	CLMSDRPYCGMDGCRGISIYSSERSVL--QSINPAEDHKECPNPKPDKAPLQKSLAPNS	528
Db	585	GSAPFCEPRSLQAHVQV 602	
Oy	529	RYSLSCMESRSHATYSW 545	
RESULT	5		
ID	062181	PRELIMINARY; PRT; 751 AA.	
AC	062181;		
DT	01-NOV-1998 (TREMBLREL, 08, CREATED)		
DT	01-NOV-1998 (TREMBLREL, 08, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBLREL, 08, LAST ANNOTATION UPDATE)		
DE	SEMAPHORIN E, PRECURSOR (SEM E).		
GN	SEMAE OR SEME.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUFAROTIA; METAQOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MORIDAE; MURINAE; MUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-NMRI;		
RX	MEDLINE; 95267431.		
RA	PUESCHEL A.W., ADAMS R.H., BETZ H.;		
RT	"Murine semaphorin D/collapsin is a member of a diverse gene family and creates domain inhibitory for axonal extension.";		
RL	NEURON 14:941-948(1995).		
CC	-I- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.		
CC	-I- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).		
CC	-I- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.		
CC	MAXIMUM EXPRESSION BETWEEN DAYS 10-12 WITH MODERATE LEVELS FROM DAY 13 UNTIL BIRTH.		
CC	-I- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.		
CC	CONTAINS ONE C2-LIKE DOMAIN.		
DR	EMBL; X85994; G854332; -		
DR	MGI; MGI:107557; SEMAE.		
DR	PFAM; PF00047; Ig; 1.		
KM	SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;		
KM	DEVELOPMENTAL PROTEIN.		
FT	SIGNAL 1 20	POTENTIAL.	
FT	CHAIN 21 751	SEMAPHORIN E.	
FT	DOMAIN 587 649	IG-LIKE C2-TYPE DOMAIN.	
SO	SEQUENCE 751 AA; 85259 MW; B28DCFE CAC32;		
Query Match	16.5%; Score 725; DB 11; Length 751;		
	Best Local Similarity 33.0%; Pred. No. 3,37e-151;		
	Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;		
Db	167	SENPNTVTSVMIINEELFSGMTI-DEMGTDALFRSLTKRMQLFTDQHSKWLSEPPFVD	225
Oy	139	PSPSPENSLVFEEDDEVSTIRKQENYNGK-IPIRFRRIIGSESLYTS--DTV-MQNPFQIK	194
Db	226	AHVDPDGPDAVAVYFFPKERLIDNNKSRITQIHSIMARICPNDTGGQRL-VNKTTFEL	284
Oy	195	ATVHQD-QAIVDDIYFFREDNDPKKPEAPLVNSRAQLCRGGQSGESSLVSXKMTFL	253
Db	285	KARLVCSVTDGDEGTHDELEDFLL-ETDNP-RTLVYGIPTSSSVKGSVAVCYHL	342
Oy	254	KAMLVCS--DA-AINKPNRLQDVFLFLPDSGQWRDTRVYGVSNPN-NY--SAVCYSL	307
Db	343	SDIQTVENGPAHKEGPNHOLISYQGRIPYPRPGTCGGAFTPMRTTKDPDDVYFIR	402
Oy	308	GIDIVFRTS-SLK-G-----YHSLPNDNRPCKLPDQ-QP-I-PTETF--QVADRHP	353
Db	403	NHPLMINSISIHRRPLIVRIGDTKYTKTIAVDVNAADG-RIVHVLFLGDRGTQVKVY	461
Oy	354	EVAQVKEVMEGPL-KTPEL-H--SKYHQAQVAHNRMQASHETFHVLLTLDRTGRIHKVE	409
Db	462	LPTNSSAGGELLELELEFKNHVYITTMELSSKKOQLYVSSNGEVSQVSLRHCHVGTAC	521

QY 410 -PEGOEHSFAFNIMEIOFFRRAAIIOTMSLDAERKLYVSSQWMEVSOVPLDCEVYGGC 468
Db 522 ADCLADPFCAMDGHCSPRYPTGKRSRRODYRHGNPLDQCGFNLKAVRNAEIVQY 581
QY 469 HGLMSDPLPCGMDQGCISITYSSE--RSVLQSIINPAEPHKECP--NPKPKAPLQKYSL 524
Db 582 GVRNNSTFELCAPKSPQASIKMLQKDK 609
QY 525 A-PSRBYLSCPMESRHATYSWR-HKEN 550

RESULT 6
ID Q13214 PRELIMINARY: PRT: 749 AA.
AC Q13214;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 96210603.
RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA ALBAHESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT cancer deletion region and demonstrate distinct expression
RT patterns";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR EMBL: U28369; G974284; -.
DR PFM: PF00047; 19; 1.
SQ SEQUENCE 749 AA; 83121 MW; 6138C5D5 CRC32;

Query Match 16.38; Score 716; DB 4; Length 749;
Best Local Similarity 33.08; Pred. No. 7,90e-149;

Matches 167; Conservative 115; Mismatches 170; Indels 54; Gaps 41;

Db 112 ECMNFVKLLAAYNRTHLACGTGAFHPTCAFEVGHRAEPPVRLDPRIGEDGKSPYD 171
QY 93 DCEYITLLER-RSEGLLACGTNARHPSG-WNLV-N-G--TVVPL-G--EM-RGIAPFS 141
Db 172 PRRAASVLVGEELYSVA-ADLMGRDFTIFRSLGQRPSTLTERPHDSRWLNPRKPVFW 230
QY 142 PDENSLVLFEGDEVYSTIRKQYNGK-IPRFRIRGSELYT-S-DIV-MQNQFIKAT- 196
Db 231 IPSENDDDKITFFRETFEVAEPALGRLSVSRVGOICNDVGGORSL-VNKWTFELKA 289
QY 197 IVHODQAYDKIYFFREDNPDKNPE-APLNVSRVQOLCRGDGEGSSLSVSKWNTFLKA 255
Db 290 RLVCYSGVEGDTHFDLOLVFLLS-SRDH-RTPLLYAVFSTSSIFOGSAVCVYSMNDV 347
QY 256 MLVCS--DAATNNFNRLQVFLPDPSCGWRDTRVYGVFSNFWN-Y--SAVCVYSLGDI 310
Db 348 RRAFLGFPAKKEGPMQWVSYQGRVYPRPGMCPKTFGTSSTKDFPDDVIOFARNHPL 407
QY 311 DK--V--F--RTSSL-K--GYHSSLPRMPRGKCLPQ-QPIP-TEFF--QVAD--R-HPE 354
Db 408 MYSVLPTG-R-PLFQVQANTFTIOIADRYAAGH-YDVLFTGTDVGYLKYISVP 464
QY 355 VAORVEPMGLKTRPLF-H--SKYHQVAVHARMQASHGEFFHLYLTDRGTIHKVVE-P 410
Db 465 KGSRPASGELLLELHVEFDSAAVTSMQISSKRHOLYVARSVAVOIALHCAHGRVCT 524
QY 411 -GEOHSFANINIEIOFFRRAAIIOTMSLDAERKLYVSSQWMEVSOVPLDCEVYGGC 469
Db 525 ECLLADPFCAMDGVACTRFOPSAKRFRRODYRNGDPSLTCGSDSRPALLEHKEVGE 584
QY 470 GCLMSRPPYCGMDQGRICISYSSERSVL-QSINPAEPHKECPNPKPKAPL-QKYSLAP 526
Db 585 GSSAFLECEPRSLQARVEMTFORAGV 610

QY 527 NSRYLSCPMESRHATYSWR-HKENV 551

RESULT 7
ID Q93018 PRELIMINARY: PRT: 750 AA.
AC Q93018;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA DANTE M., WAMSLEY P.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U73167; G2880035; -.
DR PFM: PF00047; 19; 1.
SQ SEQUENCE 750 AA; 83034 MW; CB87E34 CRC32;

Query Match 16.38; Score 716; DB 4; Length 750;
Best Local Similarity 33.08; Pred. No. 7,90e-149;

Matches 167; Conservative 115; Mismatches 170; Indels 54; Gaps 41;

Db 113 ECMNFVKLLAAYNRTHLACGTGAFHPTCAFEVGHRAEPPVRLDPRIGEDGKSPYD 172
QY 93 DCEYITLLER-RSEGLLACGTNARHPSG-WNLV-N-G--TVVPL-G--EM-RGIAPFS 141
Db 173 PRRAASVLVGEELYSVA-ADLMGRDFTIFRSLGQRPSTLTERPHDSRWLNPRKPVFW 231
QY 142 PDENSLVLFEGDEVYSTIRKQYNGK-IPRFRIRGSELYT-S-DIV-MQNQFIKAT- 196
Db 232 IPSENDDDKITFFRETFEVAEPALGRLSVSRVGOICNDVGGORSL-VNKWTFELKA 230
QY 197 IVHODQAYDKIYFFREDNPDKNPE-APLNVSRVQOLCRGDGEGSSLSVSKWNTFLKA 255
Db 291 RLVCYSGVEGDTHFDLOLVFLLS-SRDH-RTPLLYAVFSTSSIFOGSAVCVYSMNDV 348
QY 256 MLVCS--DAATNNFNRLQVFLPDPSCGWRDTRVYGVFSNFWN-Y--SAVCVYSLGDI 310
Db 349 RRAFLGFPAKKEGPMQWVSYQGRVYPRPGMCPKTFGTSSTKDFPDDVIOFARNHPL 408
QY 311 DK--V--F--RTSSL-K--GYHSSLPRMPRGKCLPQ-QPIP-TEFF--QVAD--R-HPE 354
Db 409 MYSVLPTG-R-PLFQVQANTFTIOIADRYAAGH-YDVLFTGTDVGYLKYISVP 465
QY 355 VAORVEPMGLKTRPLF-H--SKYHQVAVHARMQASHGEFFHLYLTDRGTIHKVVE-P 410
Db 466 KGSRPASGELLLELHVEFDSAAVTSMQISSKRHOLYVARSVAVOIALHCAHGRVCT 525
QY 411 -GEOHSFANINIEIOFFRRAAIIOTMSLDAERKLYVSSQWMEVSOVPLDCEVYGGC 469
Db 526 ECLLADPFCAMDGVACTRFOPSAKRFRRODYRNGDPSLTCGSDSRPALLEHKEVGE 585
QY 470 GCLMSRPPYCGMDQGRICISYSSERSVL-QSINPAEPHKECPNPKPKAPL-QKYSLAP 526
Db 586 GSSAFLECEPRSLQARVEMTFORAGV 611
QY 527 NSRYLSCPMESRHATYSWR-HKENV 551

RESULT 8
ID Q99985 PRELIMINARY: PRT: 751 AA.
AC Q99985;

D	346	DIIVNENPFNAHKGNPNQWMPFGSKMPPRGPCPGFTFPBSKSKNDYDVEYINMRS	405
Q	309	DIIVNENPFNAHKGNPNQWMPFGSKMPPRGPCPGFTFPBSKSKNDYDVEYINMRS	351
D	406	HPLMQAVYPLQRPPLVVRT-GAPYRLTITVAVDQVSADG-RYEVLTGDRGTQKVIY	463
Q	352	HPYVAQNRPEPG--PLKTRPLFSKXTHQKAVNHMQASHGTFHVLITLDRGITHKVE	409
D	464	LPKDDQEMELMLEEVEVEFKDPAPKVTWTISSKROQLYASAVGVTLSLHRCQAYAAC	523
Q	410	PGEOEHSEFAFINIMEIQFRRAAAIQTWLSDAERKKLYVSSQWVEVQYPLDLCEVGGC	468
D	524	ADCLLAPRYCAMGQACSRATSSAKRSRQDVRHGNPTQCGGFSNANKNAVESQY	583
Q	469	HGCLMSRPYCGMOGRIS-IYSER-SYLOSINPAEPHKEP--NPKPDKAPLQKVS	524
D	584	GVAGSAFLFCOPRSPQATVKM 605	
Q	525	A-PNSRYLSPMESRATYSW 545	
RESULT	11		
ID	088633	PRELIMINARY:	PRT: 754 AA.
AC	088633		
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DE	SEMAPHORIN IV ISOFORM A.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
NC	SCUROGNATHI; MORIDAE; MORINAE; MUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	ECKHARDT F., MEYERHANS A.;		
RT	"Molecular cloning and expression pattern of a murine semaphorin		
RL	homologous to h-sema iv."		
SR	SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.		
DR	EMBL; AF080091; G3377768; -		
SO	SEQUENCE 754 AA; 85133 MW; EC05BDF3 CRC32;		
Query Match	15.5%;	Score 682;	DB 11; Length 754;
Best Local Similarity	31.8%;	Pred. No. 6.75e-140;	
Matches	141;	Conservative 119;	Mismatches 140; Indels 43; Gaps 31.
D	171	PYDPLADTASALINEELAGYVI-DEMGTDAIFRTLGOTAMRTDOYNSKRLNDPSFIH	229
Q	139	PFSDENDENVLEFGEDEVSTIRKQENYCK-IPRRIRIGSELYTS--DIY-MQNPQFIK	194
D	230	ALPIPDSARNDDKLYFFERERSAE-AFQNPAYVARIGRICLNDGQHCL-VNKMSTFL	287
Q	195	ATIVHOD-QAVDDQIYFFREDNDKNEPDLANSRAVQLCRDQGESSLVSKWMTFL	253
D	288	KARLYVCSFPGDGLIETHDELQDFV-QQITDI-RNPITYAVFTSSGVSFEGSAVCYISM	345
Q	254	KAMLYVCS--DAANKNENRLLQDVFLLPDPGQMRDTRVYGVFNSP-WNY--SAVCYSL	307
D	346	ADIRNVFNGPRANKEGPRYQWMPSPGKMPYRPGTCPGGTFTPMKSTKQYDVEINMR	405
Q	308	GDIDKVFYTS-SLK-G-YH-----SS-LPNRPFGKCLPDC-QP-I-PTEIF-QVAD--R	351
D	406	THPLMYQAVYPLQRPPLVVRT-GAPYRLTITVAVDQVSADG-RYEVLTGDRGTQKVIY	463
Q	352	HPYVAQNRPEPG--PLKTRPLFSKXTHQKAVNHMQASHGTFHVLITLDRGITHKVE	409
D	464	LPKDDQEMELMLEEVEVEFKDPAPKVTWTISSKROQLYASAVGVTLSLHRCQAYAAC	523
Q	409	E-PGEOEHSEFAFINIMEIQFRRAAAIQTWLSDAERKKLYVSSQWVEVQYPLDLCEVGGC	468
D	524	CADCLLADPRYCAMGQACSRATSSAKRSRQDVRHGNPTQCGGFSNANKNAVESQY	583
Q	468	HGCLMSRPYCGMOGRIS-IYSER-SYLOSINPAEPHKEP--NPKPDKAPLQKVS	524
D	584	YGVAGSAFLFCOPRSPQATVKM 606	

RESULT	ID	PRELIMINARY:	PRT:	772 AA.
AC	090607			
DT	01-NOV-1996 (TREMBLREL, 01, CREATED)			
DT	01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL, 08, LAST ANNOTATION UPDATE)			
DE	COLLAPSEIN.			
OS	GALLUS GALLUS (CHICKEN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSUURA; AVES;			
OC	PHAGNOMYXINAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-BRAIN;			
RX	MEDLINE: 94006554.			
RA	LUO Y., RAIBLE D., RAPER J.A.;			
RT	'Collapsein: a protein in brain that induces the collapse and			
RT	paralysis of neuronal growth cones.';			
RL	Cell 75:217-227(1993).			
DR	EMBL: U02528; G410079; -.			
DR	PRAM: PF00047; 19; 1			
SO	SEQUENCE 772 AA; 88867 MW; BC6CB98 CRC32;			
Query Match	15.5%; Score 681; DB 13; Length 772;			
Best Local Similarity	31.5%; Pred. No. 1,24e-139;			
Matches	146; Conservative 119; Mismatches 150; Indels 48; Gaps 38			
D	166 RGSFDPDKLTASLVGDELVSSTA-ADFGCRDFAIFRTLGHHNPIRTQHSRWLNDP 224			
Q	135 RGAFAFSPDENSLVLEEGDEVSTIRKOEYNGK-IPRRRIIGSESLYTS--DTV-MONP 190			
D	225 RFIASHLIPESNDPEEDKTYFFERFENALIDGEGTKATPHARIGCICKNDPFGHSL-VNKK 283			
Q	191 OFIATVATVHO-DDAYDDKITYFFREDNPDKNEAPLNVSRVAOLCRDGDGESSLSVSKW 249			
D	284 TTELKARLTCSVGPNGIDTHEDELDQDFLM-NSKDP-KNPIYGVATYTSNIFKGSAYC 341			
Q	250 NTEFLKALVNS--DA-ATMKNNRRLQDVFLLRDPSCGMQRDRYRYGVYSNWN-Y--SAYC 303			
D	342 MYSTMDVRRVFLGPRVYAHNRDGPYQWVYQGRVYPRPQETCSPTFGGFDSTKDLPDEVIT 401			
Q	304 VYSIGDIDKIFRTS-SLK-G-YH--S-S-LPNRPQKCLPQ-QPI-PTETFF-QV-A 349			
D	402 FASHSHAMKYPRVPRINS-PRMIKIKDVYQGTQIYVDRVRADEQ-QIDVAFITDIDGIVL 459			
Q	350 -DR-HPEVAQVREPMGRPLTLPLEHSHK--YHOKVAVHNKMQSHETNVLVLTITDGTHT 405			
D	460 KVASIPRETHELEEVLEEMTVFREPTVISAAMKISTKOOOLYGSATVSQLPHRCVY 519			
Q	406 KYVE-PEQDGHSAFNPIME-IQPFRAALIQTMSIDAEKRRLYSSQWEVSQVPLDCEY 463			
D	520 YGRKACBCLLARDPYCAMPDSSCSRYFTAKRRTRRQDRIKNGDPLTGHSDLDHNDPSCQ 579			
Q	464 YGGGCHGCLMSRDPYCGMDQGRICISITYSE-RSYL-QSINABENHKCSRN-PRKRAP-- 518			
D	580 TLEBKIIITYGVENSITFLECSPRSQALYY-WQFOKQNDKDKVE 621			
Q	519 -LQ-KVSLA-PNSRYVLSL-C-PRESHATYTSWR-HKEVNEQSE 556			
RESULT	13			
ID	088632	PRELIMINARY:	PRT:	765 AA.
AC	088632;			
DT	01-NOV-1998 (TREMBLREL, 08, CREATED)			
DT	01-NOV-1998 (TREMBLREL, 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL, 08, LAST ANNOTATION UPDATE)			
DE	SEMAHOKIN IV ISOFORM B.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.			

[illegible]

Search completed: Thu Jul 8 19:16:01 1999
Job time : 81 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:20:58 1999; Maspar time 29.65 Seconds

Tabular output not generated. 416.785 Million cell updates/sec

Title: >US-09-041-236-2
(26-606) From US09041236.pep (33 of 45)

Perfect Score: 4344

Sequence: 1 MKGHVGQDRVDFGQTEPHTV.....SYFREAQHWQLLPEDQINAE 581

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 2126608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.421; Variance 154.264; scale 0.236

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	661	15.2	771	13	Human semaphorin III	8.10e-51
2	620	14.3	477	13	Human collapsin.	7.99e-47
3	608	14.0	775	33	Human semaphorin	1.17e-45
4	606	14.0	861	22	Mouse CD100 antigen.	1.83e-45
5	599	13.8	861	32	Human semaphorin.	8.76e-45
6	544	12.5	862	32	Human semaphorin.	1.85e-39
7	512	11.8	776	32	Rat semaphorin W.	2.24e-36
8	463	10.7	441	13	Vaccinia virus semaph	1.11e-31
9	459	10.6	730	13	Grasshopper semaphorin	2.67e-31
10	428	9.9	587	33	Human semaphorin W.	2.38e-28
11	390	9.0	974	33	Human secreted protei	9.32e-25
12	378	8.7	712	13	Tribolium semaphorin	1.35e-23
13	367	8.4	650	13	Drosophila semaphorin	1.35e-22
14	362	8.3	724	13	Drosophila semaphorin	3.96e-22
15	352	8.1	930	32	Human semaphorin Y.	3.40e-21
16	325	7.5	888	25	Human semaphorin Z.	1.10e-18

17	316	7.3	887	25	W19856	Rat semaphorin Z.	7.44e-18
18	296	6.8	929	32	W57259	Rat semaphorin Y.	5.12e-16
19	157	3.6	122	13	R13185	Varicella major virus s	8.25e-04
20	99	2.3	591	4	R23006	Protein transcribed f	2.54e+01
21	98	2.3	832	33	W61092	Taq DNA polymerase I	2.99e+01
22	101	2.3	2329	25	W25038	Partial BRCA2 cancer	1.93e+01
23	96	2.2	832	33	W61090	Taq DNA polymerase I	4.14e+01
24	91	2.1	478	1	R04881	Recombinant elastase I	9.18e+01
25	91	2.1	598	9	R48631	Sequence of nuclear r	9.18e+01
26	90	2.1	638	5	R03924	E. coli HSP (dnak).	1.07e+02
27	92	2.1	832	33	W61091	Taq DNA polymerase I	7.84e+01
28	92	2.1	854	30	W56309	Class II S-receptor Ki	7.84e+01
29	92	2.1	855	6	R29815	S receptor kinase pro	7.84e+01
30	90	2.1	1242	13	W52287	Rattus norvegicus cdo	1.07e+02
31	93	2.1	4572	30	W52845	A. mediterranei rifam	6.69e+02
32	89	2.0	534	25	W25031	Partial BRCA2 cancer	1.26e+02
33	89	2.0	554	13	R66209	Novel thermostable DN	1.26e+02
34	89	2.0	680	4	R23143	Mutant thermostable D	1.26e+02
35	89	2.0	832	13	R76693	DNA-polymerase (F3),	1.26e+02
36	89	2.0	832	13	P90556	Purified native therm	1.26e+02
37	89	2.0	832	33	W61087	Taq DNA polymerase I	1.26e+02
38	89	2.0	832	33	W61088	Taq DNA polymerase I	1.26e+02
39	89	2.0	832	13	R76690	Taq DNA-polymerase RE	1.26e+02
40	89	2.0	832	33	W61089	Taq DNA polymerase I	1.26e+02
41	89	2.0	833	36	W59942	Amino acid sequence o	1.26e+02
42	89	2.0	833	27	W24211	Cleavage DN nuclease.	1.26e+02
43	89	2.0	3418	26	W19211	Cleavage DA nuclease.	1.26e+02
44	89	2.0	3418	26	W19211	Human breast cancer s	1.26e+02
45	89	2.0	3418	26	W23287	Human breast and ovar	1.26e+02

ALIGNMENTS

RESULT	1	R71380 standard: Protein: 771 AA.
ID	AC	R71380;
DT	21-NOV-1995	(first entry)
DE	Human semaphorin III protein.	
KW	Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;	
KW	Varicella major virus; smallpox; semaphorin receptor binding activity;	
KW	modulation; nerve cell growth; immune response; viral pathogenesis;	
KW	neurological disease; neuro-regeneration; oncological infection.	
OS	Homo sapiens.	
PN	W09507706-A.	
PD	23-MAR-1995.	
PF	13-SEP-1994; U10151.	
PR	13-SEP-1993; US-121713.	
PA	(REGC) UNIV CALIFORNIA.	
PI	Bentley DR, Goodman CS, Kolodkin AL, Matthes D;	
PI	O'Connor T;	
DR	WPI: 95-131177/17.	
DR	N-PSDB: 087442.	
PT	New class of semaphorin peptide(s) and polypeptide(s) - are	
PT	potent modulators of nerve cell growth and regeneration	
PS	Example 2; Page 60-63; 101pp; English.	
CC	The sequence of the human semaphorin III protein. The proteins	
CC	encoded by the grasshopper semaphorin I (087441), human semaphorin III,	
CC	vaccinia virus semaphorin IV (087443), Drosophila semaphorin I and II	
CC	(087444-5), Tribolium semaphorin I (087446) or varicella major (smallpox)	
CC	virus semaphorin IV (087447) genes were used to generate a series of	
CC	peptides (R70370-R70418), which retain semaphorin receptor binding	
CC	activity. The semaphorin derived or semaphorin receptor derived peptides	
CC	are potent modulators of nerve cell growth, immune responsiveness and	
CC	viral pathogenesis. They can be used in diagnosis and treatment of	
CC	neurological disease and neuro-regeneration, immune modulation and	
CC	diagnosis and treatment of viral and oncological infection and diseases.	
SO	Sequence 771 AA;	
Query Match	15.2%; Score 661; DB 13; Length 771;	
Best Local Similarity	30.3%; Pred. No. 8.10e-51;	
Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;		
Db	166 rkykypklltasllldgelysgta-adfmgcdfaifrlghhprtqghdswindp 224	

Oy	135	RGYAFSPDENSILVFEDEDEYVSTIRKOEYIGK-IPRFRRIRGSESLYTS-DTV-MONP	150
Db	225	kfishelisesinpeddkvfyifrenalidgeshgkatharigickndfgghrsyl-vnw	283
Oy	191	QFIKATIVHQ-DQAVDDKIYFFREDNPDKNPEARLVNRYAOLCRRDQOGESLSYSKV	249
Db	284	ttfllkarlcsvporngidthfdeqdfim-nfkdpr-kirpyuvgvfetsnifksavc	341
Oy	250	NTEFLKAMVLVS--DA-ATNKNFNRLQDFELLPDRDSGOMRDTRYGVGSNPMW-Y-SAVC	303
Db	342	mysmdvrrvrfgrahhdgprnywvrgqvgvrrprrprrpsrsktfggfdfscktdprddvit	401
Oy	304	VYSLGDDIKVRTS-SLK-G--YH--S--S-LPRPRRCKLPRD-QPI-PRTEF--OV-A	349
Db	402	farsipamyrbvfmmnprviktdvngftqivdvtdvaeedqg-ydvmfigtvgtvxl	450
Oy	350	-DR-HPEVAQVERBGRKLTPRL-FHSKXHYQKVAUHNMQASHGEBTHVLVLTTRDGRITIK	406
Db	461	vvsjlpketwyldeevllemvtvfereptaisemelstkgqqluygstgvaqlrlnchdly	520
Oy	407	VVE-PGEDEHSFAFINIME-IOPFRRAAIQMSIDASRRKRLYVSOWMEVQOVRPLDCEVY	464
Db	521	gkacaeccrlardpycawgsagscsyufbrakttrtrgqilngdrplthcsdlihhnbnhbsp	580
Oy	465	GGCGGCLMSMDPFCGMDQGCISITYSSE-RSVL-QSINRAEPKKECPNKRPRDK----	AP 518
Db	581	eerililyvensstfiescpksqralvywqfgrnneerkeelirvddhllirtcdqlllrslq	640
Oy	519	LQKVELA-PNSRYVLSC-PMESRHATY-SW-RHKENV-EGQCEGCHQ-SPNCILFIENTLT	572
Db	641	qkdsngnylchavengfiq	658
Oy	573	AOQYGHYCEAOEGSYFR	590

ID	RESULT	2	AC	R7A175	standard; Protein; 477 AA.
DT	01-NOV-1995	(first entry)	DE	Human collapsin.	
OS	Collapsin; antibody; therapy.		CS	Homo sapiens.	
FT	Key	Location/Qualifiers	FT	binding_site	9..19
FT	binding_site	/note= "antibody binding site"	FT	binding_site	51..65
FT	binding_site	/note= "antibody binding site"	PN	US5416197-A.	
PN	16-MAY-1995.		PF	15-OCT-1993; 136922.	
PA	15-OCT-1993; US-136922.		PR	(USPE-) UNIV PENNSYLVANIA.	
PI	Luo Y, Raper UA.		DR	WPI; 95-1934/8/25.	
DR	N-PSDB: Q92331.		PT	New antibody to human collapsin - used to inhibit the activity of	
PT	collapsein. to induce neurite out-growth and to treat individuals with		PT	nerve damage.	
PS	Claim 2; Columns 15-18; 11pp. English.		CC	An antibody capable of specifically binding at least a portion of	
CC	the collapsein protein can be used to purify human collapsin and		CC	to inhibit the activity of the protein. It can be used to induce	
CC	neurite outgrowth by neuronal cells and to treat individuals		CC	suffering from nerve damage.	
SO	Sequence	477 AA;			
Query Match	14.3%;	Score 620;	DB 13;	Length 477;	
Best Local Similarity	31.2%;	Pred. No. 7,99e-47;			
Matches	134;	Conservative 110;	Mismatches 145;	Indels 41;	Gaps 32;

```

Oy 198 VHO0A0YDKK1YYFFREDNPDKNPEADPLN7SRVA0LCRG0GGESSLSVSKWMTFLKML 257
Db 60 lcsvprgnlthldelqdvflm-nfkdp-knpvygvftcsnlfkysavcmymadvc 117
Oy 258 VCS--DA-ATKKNRNL0ADVFLLRDPBGG0MRDTRFVYVFSNPMN-Y--SAYCYSLGDID 311
Db 118 rvflgryahrcdprnqygvwpygvgrvprpdpccpsktlfgdstckdipddvltafarshpm 177
Oy 312 KVFRRS-SLK-G--YH--S--S-LPNRRPKCLPDC-QPI-PIETF--QV-A-DR-HPEV 355
Db 178 ynpvfpmnnpriaviklctdnygftglvvdvdaedqg-ydvmfigtcdvgtlkwvsipket 236
Oy 356 AQRERPMGRPLKTRPL-FHSKNHYQAVANHRQASNGERFHVLLYLTDRGTHKNVLE-PGEQ 413
Db 237 wydlleevlleemtyfcreptalsamejstkbqqlylgsttagvagi1rlhrcdiygkasaecc 296
Oy 414 EHSFANFIME-IGPFRRRAAIQTMSLDAERKKLYVSQMEVSEVPRDLCEVGGCGHCL 472
Db 297 lardpycawdgsacsyrfptakrttrtqgdltngdplhscdlhhdnhghspeerllygv 356
Oy 473 MSRPBRCMDQGRCSISYSS-RSVL-QSINPAVEPHKECENPKRDK---APLOKSLA- 525
Db 357 ennsrflecspsksgraluyngqfgrmeerkeelrvdhhlttrdggl1lrslgqdsnyl 416
Oy 526 PNSRYLYLSC-PMESRHAUY-SW-RHKENV-EQSCBPCHQ-SPKCIETIENTLTAQOYGHVF 580
Db 417 chavengfiq 426
Oy 581 CEADGGSYFR 590

```

```

RESULT      3
ID          ID      W63748 standard: Protein: 775 AA.
AC          AC      W63748;
DT          DT      01-OCT-1998 (first entry)
DE          DE      Human semaphorin.
KW          KW      Human; semaphorin; atiposis; nerve disease; nerve growth inhibitor;
KW          KW      neurological disease; atopic skin inflammation; autoimmune disease;
KW          KW      pain.
OS          OS      Homo sapiens.
PN          PN      W09822504-A1.
PD          PD      28-MAY-1998.
PE          PE      12-NOV-1997; J04111.
PR          PR      15-NOV-1996; JP-321068.
PA          PA      (SUMU ) SUMITOMO PHARM CO LTD.
PI          PI      Furuyama T, Inagaki S;
DR          DR      WPI: 98-312416/27.
N-PSDB:    N-PSDB: V35367.
PT          PT      Gene encoding new semaphorin nerve growth inhibitor - useful in
PT          PT      diagnosis, treatment and study of neurological diseases
PS          PS      Claim 1; Page 33-37; 49pp; Japanese.
CC          CC      The present sequence represents human semaphorin, a nerve growth
CC          CC      inhibitor. The semaphorin protein, and gene encoding the protein,
CC          CC      and their derivatives, are used in the diagnosis, treatment and
CC          CC      study of neurological disorders such as atopic skin inflammation,
CC          CC      autoimmune diseases and pain.
SQ          SQ      Sequence 775 AA;

Query Match      14.0%; Score 608; DB 33; Length 775;
Best Local Similarity 32.1%; Pred. No. 1,17e-45;
Matches 125; Conservative 98; Mismatches 130; Indels 36; Gaps 28.

Db 167 rgrcpfpnussfvtlvgnelafaglys-dywgtrdsalfrsmgk1ghlrltchdderillkcp 225
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 135 RGVAPFSPBDENSLVLFEFGDEVEYSTRKQENGRK-IPFRRI-R-GE-SELYTSDTWQNP 180
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 226 kfyasymphdmedrddhkmyfftekleaeeannahctyrtgrvrlcvmndmgqgrll-vnkx 284
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 191 OFIKATIVHQDQAYD-KIYIFREDNPDKNPEAPLVNSVRAQICRQDQCGSSLSKMW 249
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 285 stflkarlivesvpmqgldcyfdeledvlllp-trdp-knpvifglfntslsflfrhvac 342
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 250 NTFELKAVSCDADATN-KN--FNFLQDVFLLPDPGSGMRDTRVYGVSPFN-N-Y-S-AVC 303
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```


Db	343	ylhmssiteafnfygryahkegepeynslsylegkvpyrprogscaaskvngskyglttkdypddai	402
Oy	304	YVSLDDIDDKVTRTS-SLK-G- YVSSLT-----PMPRGKCLP-D-QDP1--PTERRF-OVAD	350
Db	403	rfarahpilmypgipkivkhkpllvltvdgklyuirlavavrveedq-ydvlfiqlvdlcivl	461
Oy	351	RHPEVAQNV-EPMGCL-KTP-LFHS--KYHQAIVAHVMAQASHETRHVLVLTTRKDTIH	405
Db	462	kvltlmgsetewemeevileejlqikdpapilismelskskrqlylgsasavayrfhncdm	521
Oy	406	KVVEGCEDEHSEFAFN-IME-IQPRRAAAIQTMSLDERRRLYVSQMEVSQVPLDCEV	463
Db	522	ygsacadcclardpycawdgjiscrrypt	550
Oy	464	YGGGCHGCLMSRNDPTCGMDQRCISITISS	492
RESULT 4			
ID	17658	standard; Protein; 861 AA.	
AC	M17658:		
DC	24-JUL-1997	(first entry)	
DE	Mouse CD100 antigen.		
KW	CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;		
KW	vaccine.		
OS	Mus sp.		
FT	Key	Location/Qualifiers	
FT	peptide	1..41	
FT	protein	/label= Sig-peptide	
FT		42..861	
FT	domain	/label= Mat-protein	
FT		42..553	
FT	domain	/label= Semaphorin_domain	
FT		554..630	
FT	domain	/label= Ig-like_domain	
FT		631..732	
FT	domain	/label= Scalk_domain	
FT		734..752	
FT	domain	/label= Transmembrane_domain	
FT		753..861	
FT	domain	/label= Cytoplasmic_domain	
FT		807..814	
FT	modified_site	/label= Phosphorylation	
FT		/note= "putative tyrosine phosphorylation site"	
FT			
FT			
FN	W09717368-A1.		
PD	15-MAY-1997.		
PR	12-NOV-1996; UI8645.		
PR	09-NOV-1995; US556422.		
PA	(DAND) DANA FARBER CANCER INST.		
PI	Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL,		
DR	WPL: 97-280982/25.		
DR	N-PSDB: T60666.		
PT	Nucleic acid molecule encoding CD100 antigen - which stimulates		
PT	leukocyte response, e.g. B cell aggregation, differentiation,		
PT	survival and T cell proliferation		
PS	Example 8: Page 86-89: 13pp: English.		
CC	Mouse CD100 antigen (M17657) is a novel leukocyte semaphorin-like		
CC	protein that stimulates a leukocyte response, including B cell		
CC	aggregation, B cell differentiation, B cell survival and/or T cell		
CC	proliferation. Its amino acid sequence was deduced from a cDNA		
CC	clone (T60666) isolated from murine T cells. Human CD100 antigen		
CC	(M17657) has also been identified. CD100 polypeptides and fusion		
CC	proteins, nucleic acids, and host cells expressing CD100 can be		
CC	utilised in diagnostic and therapeutic methods involving modulation		
CC	of B and T cell responses, neuron axonal growth and immune cell-		
CC	nerve cell interaction.		
CC	Sequence 861 AA:		
CC			

Query Match	14.0%;	Score 606;	DB 22;	Length 861;
Best Local Similarity	32.4%;	Pred. No. 1.83e-45;		
Matches	143;	Conservative 106;	Mismatches 145;	Indels 47;
				Gaps 31.

Db 107 ecInyirvIqplststlyvcgtnafqptcdhInItsfkflqksedgkqrcpfdpahsyts 166

QY 93 DCENITTLER-RSGLLACCTMRHPSG--WNLVNCTVPLGEM-RGYAPFSPDENSLV 148
167 ymvggellsgts-ynflgsepilismshspilrteyalpwlnepsfyfadvgkspdpge 225
149 LFEGEVYSTIRKOEYNGKIPRFRIRIGESLTYSDTV--MNPQFIKATIVHQ--D-Q 202
Db 226 geddkvvyffifvevsvevefvfklnipravarvckdggglitlg-kkustfkarilockp 284
QY 203 AADDDIYFFEFRRDNDPKKNEAPLWVSRAQCRDGOESLSVSKNNTFLKMLVCSDA 262
Db 285 deglyfnllgdfvflrap-gl-kepvfayvtpqlnnvglsavcaeylatveavfsgky 342
QY 263 ATKNFENLQDVFLLPPDSGGMROTRRYGVSPNWNY---SAYCVISLGIDDKVF-R--- 315
Db 343 mgstatvegshckwryngpvpbtpipgacidsaaraanytslnlpckltlgfvkdhplmd 402
QY 316 --TSSL-----K-G-YHSLLENPRPGKCL--PDQPIP-TETFGVADRHPE-V-AQR-VE- 360
Db 403 svtpdlnprklkkdvnytgvlvartgaldgltcfymfistdrgalbkavllckevn--v 460
QY 361 PGPSPKLT-P-LPHSKYHQAQVAVHMQASHETHTVLYLTTRDTIKHKVPEGEDEISFA 418
Db 461 --leetqgfrfsepylltlleskgrkffvagsnsgvvgaplatfecghs--cedvliard 517
QY 419 ENIMIQGFRRRAAQITMSLAER-RKL-VYSSQMEVSQVPLDCEYVGGCCHCLMSRD 476
Db 518 pycawspaikacvtllhgeas 558
QY 477 PYCGWDQG-R-CISYSSERS 485

RESULT 5
ID W58540 standard; Protein; 861 AA.
AC W58540;
DT 02-SEP-1998 (first entry)
DE Human semaphorin.
KW Human semaphorin; diagnosis: nervous disease; immune disease;
KW nerve extension inhibiting activity.
OS Homo sapiens.
PN J10153490-A.
PD 16-JUN-1998.
PF 27-NOV-1996; 332900.
PR 27-NOV-1996; JP-332900.
PA (SUDU) SUMITOMO SEIYAKU KK.
DR WPI: 98-381044/34.
DR N-PSDB: V31121.
PT New human semaforin gene - useful in the diagnosis of nervous system
PT and immune disorders
PS Claim 1: Page 10-12; 15pp; Japanese.
CC The present sequence represents human semaphorin (translated from the
CC Japanese specification as semaforin). Semaphorin has nerve extension
CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
CC treatment and researches on nervous diseases and immune diseases.
SQ Sequence 861 AA;

Query Match	13.8%	Score 599;	DB 32;	Length 861;
Best Local Similarity	32.4%;	Pred. No. 8.76e-45;		
Matches: 133; Conservative	105;	Mismatches 146;	Indels 47;	Gaps 31;

Dd 107 ecnmyrrvqlplstslsyvcgtatfgptcdlnltsfkfkgseadgkrcpfpahys 166
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 93 DCENETTLER-RSEGLIACCTNARHSC--WNLVGTVPPLGEM-NGYAPFSDENSLV 148

Dd 167 ymvgelysgcs-ynfligsepilsmshaplrtegiypwlnepsfvaavikspdgpe 225
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 149 LFEGEVESTTRKOEYNKNIPRRIRIGESLEYLSTDV--MQNQPKFAITVAIQ-D-Q 2022

Dd 226 geddkvvfffevsveyefvkllmpvarckddggdltlq-kkwtsfkariicskp 284
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 203 AYDDKITFFEFEDNDKPKEPPLAVNSRYAQLCRDDGGESSLSYSKNNTFLKMLVCSDA 2622

Dd 285 dsqlyfnllqgvfvrlrap-qf-kcpvfayftvpjnnvjisavcaaylatvaeafsqky 3424

QY : |||||:| |::|:| | ||||:| |
263 ATKNNFNRLQDFELLDPBEGQMRDTRVYGVFSNPMWY---SAVCYSLGIDIKVF-R-- 315
Db 343 mgsatvegshkwtvryngprpfpagacidsaaraanytslnlpdkltqfvdhplmd 402
OY :|||:| | ||||:| |::|:| | |::|:| | |::|:| |
316 ---TSSL-----K-G-YHSSLNPRPGKCL-PDQOPRP-TETFOVADRHPE-V-AQR-VE- 360
Db 403 svtrpIdnprlkkdvnyqtlvdrctqaldgtfvydmfistdrgalrhkavlltkvvh--v 460
OY :|:|:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
361 PMGPLKT-P-LFHSKHYHKQKVAVHRMQASHGETFHVLYLTDRGTIHKVVEPGEQHSFA 418
Db 461 --leetqldfrdepylltllsskkgtrkfyagsgnsvgavaplafeekhs-ceedcvlar 517
OY :|:|:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
419 FNIMEIQPFRRAAIQTMSLDAER-RKL-YVSSQMEVSOVPLDCEVYGGCGHCLMSRD 476
Db 518 pycawspalkacvtlhqgeas 538
OY :|||:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
477 PYCGMDQG-R-CISITYSERS 495

RESULT 6
ID M17657 standard; Protein: 862 AA.
AC M17657:
DT 24-JUL-1997 (first entry)
DE Human CD100 antigen.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
KW vaccine.
OS Homo sapiens.
FH Key
FT peptide
FT 1..41 location/Qualifiers
FT /label: Sig_peptide
FT 42..862
FT /label: Mat_protein
FT 42..553
FT /label: Semaphorin_domain
FT 554..630
FT /label: Ig-like_domain
FT 631..733
FT /label: stalk_domain
FT 735..752
FT /label: Transmembrane_domain
FT 753..862
FT /label: Cytoplasmic_domain
FT domain
FT modified_site
FT 808..815
FT /label: Phosphorylation
FT /label: "putative tyrosine phosphorylation site"
FN M09717368-A1.
PD 15-NOV-1997
PR 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND) DANA FARBEN CANCER INST.
PI Bousiolis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB: T60665.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation.
PS Claim 7; Page 70-72; 135pp; English.
CC Human CD100 antigen (M17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60665) isolated by COS cell expression cloning. CD100
CC polypeptides and fusion proteins, nucleic acids, and host cells
CC expressing CD100 can be utilised in diagnostic and therapeutic
CC methods involving modulation of B and T cell responses, neuron
CC axonal growth and immune cell-nerve cell interaction.
SQ Sequence 862 AA;
Query Match 12.5%; Score 544; DB 22; Length 862;
Best Local Similarity 29.3%; Pred. No. 1,85e-39;
Matches 157; Conservative 134; Mismatches 189; Indels 55; Gaps 39;

Db 107 eclnytrvlqplsaclslyvcgtnaifgacdhlnltsfiflqknedgkrcpfdpahsyts 166
OY :|||:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
93 DCENYITLLERSE-GLACGTMARHPCMWLVNQTVPPLG--EM-RCYAFSPDEN-SL 147
Db 167 ymvvdg-elysgts-ynffgsepilstrnsshaplregeaplylnegsfyfadviyspdp 224
OY :|:|:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
148 VLEBGEVYSTIRKOEYMGKIPRFRIRIGESLEYSDTV--MONTQFKAITVHO--D-- 201
Db 225 dgedrvyffftvevvevfvrlvlprrlarcvkgdggllctlg-kkwtcflkarllcsr 283
OY :|||:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
202 QAYDDKIYFFREDPDKNPEAPLVNSRAQLCRDQGESLSYSKNTFLKALVCSO 261
Db 264 pdsqglvfnvldvflxsp-gl-kvrvfyallftqglnnvglsavcaynlstaeefshqk 341
OY :|:|:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
262 AATNKNFNRLQDFELLDPBEGQMRDTRVYGVFSNPMWY---SAVCYSLGIDIKVF--R 315
Db 342 ymgsatvegshkwtvryngprpfpagacidsaaraanytslnlpdkltqfvdhplmd 401
OY :|:|:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
316 ---TS---S-LK-G-YHSSLNPRPGKCL-PDQOPRP-TETFOVADRHPE-V-AQR-VE 360
Db 402 dsvtrpIdnprlkkdvnyqtlvdrctqaldgtvdydmfvsctdrgalrhkal--s-lehav 458
OY :|:|:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
361 -PMGPLKT-P-LFHSKHYHKQKVAVHRMQASHGETFHVLYLTDRGTIHKVVEPGEQHSF 417
Db 459 hl-leetqldfepvqtlfllsskkgtrkfyagsgnsvgavaplafeckhgt-ceedcvlar 516
OY :|:|:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
418 AFNIMEIQPFRRAAIQTMSLDAER-RKL-YVSSQMEVSOVPLDCEVYGGCGHCLMSR 475
Db 517 dpycawspalkacvtlhqgeasvdrfslqemsgdasvcd-ks-ygsyqhffhkgtrcl 574
OY :|||:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
476 DPYCGMDQ--GRCSITYSERSVLOSINP-AEPHKECNMPDRAKPLQKVLAPNSRYLL 532
OY :|:|:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
533 SCPMESRHAITSMRKKEVQSCPE--GHQSPNCLTFENLTAQOYGHYFCEADQ 585

RESULT 7
ID W51313 standard; Protein: 776 AA.
AC W51313:
DT 08-SEP-1998 (first entry)
DE Rat semaphorin W.
KW Rat; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW Immunosuppressant; gene therapy; diagnosis; research reagent.
OS Rattus norvegicus.
PN M09815628-A1.
PD 16-APR-1998
PE 03-OCT-1997; J03549
PR 09-OCT-1996; JP-287636.
PA (SOWU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-PSDB: V07279, V07280.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Claim 1; Page 60-64; 90pp; Japanese.
CC The present sequence represents rat semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 776 AA;
Query Match 11.8%; Score 512; DB 32; Length 776;
Best Local Similarity 27.1%; Pred. No. 2.24e-36;
Matches 152; Conservative 137; Mismatches 218; Indels 53; Gaps 37;
Db 61 rfaashytnysallvdpshltlygardsifaltlpfsgcprtdmwpv-ethnqnork 119
OY :|||:| | |::|:| | |::|:| | |::|:| | |::|:| | |::|:| |
34 RYDFGQTEPHVTLFHEPSSVWVGGRGKY-L-FDPEEGKNASV-RVYNIGSTRGSC 89

QY 83 STKGC-L-DKR--DCENVITLLERRSEG-LIACGTNAHPSGMN-LVNGT-VVPLG-E 133
Db 152 grglefqpdpdnstaiyseagqlsatv-afesqtdp-li-yrg-p-ljteresdl-k-qln 204
QY 134 MGVAFPEPDESLVLFEGDEVTIRKQEVNGKIPRRIRIGSESELTSDIVMGNPQFI 193
Db 205 apufvn-tmeyndfiffifirebaveylncgkaly-srvarckhdkgg-phggagdrwtsf 261
QY 194 KATIHQOQAVDDKIFYEFREDNPD-KNPEAPLNVSRVAQLCRDQGEGESSLSVKMNTF 252
Db 262 lksrlncsvpdypfyfyeigstsliegnvgg-vekljyvtftpnslgsavcafs 320
QY 253 LKAMVVCSDAATNK-NEVRLODF-LLPDP-SGQMRDTRVYGVFSNPMN-Y--SAVCYS 306
Db 321 mksllesdpgfkeqetumswlavsikvpeprpgcvsndrsltdpvsntfvkshtlmd 380
QY 307 L-G--D-IDKVFRT--SLKG-YHS--SL--PVRPRCKCLPDDQPIPTERFQVADNHPXA 356
Db 381 eavpafit-rp1lrlisqyrfkclavdqvrtpdqkaydvlfigtdgkvikalnassaf 439
QY 357 QREVMGSLKTPLFH-S-KYHOKVAV-HRMQASHGETFHVLYLTDRGTIRKVEPEG-- 411
Db 440 dsdtdvsvleelqvlppgyvvpknllyvrmdgdsklvvsddelaiklhrsckit 499
QY 412 EOEHSFAFNIME-IQFERRAAIQTM--SLDAERRKLYSSQWBEVSQVPLDLCEVYG-G 466
Db 500 ncrecvsiqdpcaawdvclktavgsppds 530
QY 467 GCHGLMRDPRYCGWQD-G-RQISYSSERS 495

RESULT 10
ID W51314 standard; Protein: 587 AA.
AC W51314:
DE 08-SEP-1998 (first entry)
DE Human semaphorin W.
KW Human; semaphorin W; nerve extension inhibitor; anti-allergic; anticancer;
KW immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PN MO9815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T.
DR WPI: 98-261015/23.
DR N-PSDB: V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC anti-allergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 9.9%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 2,38e-28;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgddeiyffitetradsyerikvprvarvcagdlggkltl-qwtflkaddllcp 115
QY 201 DOAYDVKIYFFREDNPNKNEAPLNVSRVAQLCRDQGEGESSLSVKMNTFLKAMLVCS 260
Db 116 gpehgassvlgdvaavlpeqlga-gtclfyglffssqwgatlsavcafrpgdlrvcvlngr 174
QY 261 DAATKKNRRLQDVFLLDPDSGQWMDTRVYGVFSNPMN--YSAVCVYSLGIDIKVFRTS 317
Db 175 frelkhdncnrglpvndvnpqprpccitltnmkllfhgsslsjldrvltlflrdhplmdpr 234

QY 318 --SLK-GYHSSLP---N----PRPKCLPDDQPIP--TEFQVADRAPE-V-AQR-VE-P 361
Db 225 vfpadghllyttidcaylrvvahrvtlsqgkeyvlylgredghllhaxvrigag-ls-vl 292
QY 362 MGPLKT-PLFHSK-YHYKAVHRMQASHGETFHVLYLTDRGTIRKVEGEDEHSFAF 419
Db 293 edlal-fpeqgvemkl-yhs-w-llygsrtevyntncgrlgs-csecillaqdpvc 347
QY 420 NIMEIQFRRRAAIIQTMSLDAERRRLYVSSQWBEVSQVPLDCEVYGGSGCHGLMSRDPYC 479
Db 348 awsfldcavahaghlvgldiesadvslcpkeperpvrfevpyataahvlpssps 407
QY 480 GMD-Q-GRCISYSSERSVLOSINPAEPHKECPMPKPKAPLQKVSILAPNSRYVLSCPME 537
Db 408 samascvwhgpgsy 421
QY 538 SRHATYSWRHKEV 551

RESULT 11
ID W64221 standard; Protein: 974 AA.
AC W64221:
DE 06-OCT-1998 (first entry)
DE Human secreted protein from clone C1145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; actinin; inhibin; haemostatic;
KW chemotaxis; chemokine; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN MO9827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GENY) GENERICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Meberg D,
PI Racie LA, Spaulding V, Treacy M.
DR WPI: 98-362774/31.
DR N-PSDB: V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone C1145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, actinin/inhibin activity, chemotactic/chemokine activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 9.0%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 9,32e-25;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdecnfnkvlknddalfvcgtnafnpgsrnkmdtlepfqdefgsgmarcpydakhan 172
QY 91 KRDCENTIT-LERKSEGLACGTNAHPSGMNLVNGTVPDLG-EMKQYA--PSPSDNS 146
Db 173 valfadgklysatv-dflaidavlyrs1-gesp--tlrtvkhaskvjkpeyfvagvdyg 228
QY 147 LVLFEGDEVSTIRKQEVNG-KIPRRIRIGSESELYTSDIVMGNPQFIKATIHQOQAVD 205
Db 229 dlylyffireiveynungkvvfrvavqyckndmgsgsqvlekqvtstfkarlncsvpgds 288
QY 206 DKIYFFREDNPNKNEAPLNVSRVAQLCRDQGEGESSLSVKMNTFLKAMLVCSDAATN 265

Db	63	svladnelystv-adefsgsopllyreplrteqydsds-lnapnfv-sstf-qgd-f---	114
Oy	148	vlFEDEDEYSTRIRKOEYNGKlPRRRIRNGESELTSOTVMONPOFIATTVHDOAOAYDK	207
Db	115	vyffefrelavfincgkily-srayrcwkwdggybhr-nrwsfksrlnosipdyr	172
Oy	208	lyFFREDNPD-KNPEAPLWNRVAOUCRCDOGCESSISVSKMNTFLKALVCSDAATNK	266
Db	173	fyfneiqeasnlivegqywsmsklllygvfntrpsnslpgsawcfaiaqldadtleqgfkeq	232
Oy	267	-NFNRLODVF-LRPDPGQWMDTRVYGVFSNPMN-Y--SNAVYSLGDI-DKV---FR--	315
Db	223	tglnsmwlpymnakpdrpyscmndsfalpbdrplnfikthslmdenvyafefsqilvrt	292
Oy	316	TS--S--LKGYSHSLPMNRPCKLPDOOPRTENFEYVADRHPVEAOQVEBMGLKPTLFH	371
Db	293	stlyfyfclaydaqiklpqgktyvgvifgydthgkllkxvnaesadsdkvtsvlieidv	352
Oy	372	SK-THYQVAVNH-RMQASHGETFHVHVLITLTDRCITHK-V-VEPEE-OEH--S--FA-FNI	421
Db	353	ltkseprlnleivrtmugdqpbkdsydgdklllyvtdsqvvaqlhrcnhdkltsceeva	412
Oy	422	M-ElQPFRRAAAIQDTMSLD-A-E-R--R-KUYVSSQNEVSGVPLDLCEYVG-GGCHGCLM	473
Db	413	lqdyrcawdkiaqkrcs	429
Oy	474	SRDPYCGWDO--GRCIS	488

RESULT	14
ID	R71383 standard; Protein; 724 AA.

DT 21-NOV-1995 (first entry)
DE *Drosophila* semaphorin II protein.
KW Semaphorin; grasshopper; human; vaccinia virus; *Drosophila*; *Tribolium*,
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS *Drosophila* sp.
PN M09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPL: 95-131177/17.
DR N-PSDB: 087445.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2: Page 79-82; 101pp: English.
CC The sequence of the *Drosophila* semaphorin II protein. The gene was
CC isolated by PCR using primers based on sequence homology between the
CC complete grasshopper and partial *Tribolium* and *Manduca* (moth) semaphorin
CC gene sequences. The products gave two different sequences, each of which
CC was used to obtain its respective complete sequence: semaphorin I
CC (087444) and II. The proteins encoded by the grasshopper semaphorin I
CC (087441) human semaphorin III (087442), vaccinia virus semaphorin IV,
CC *Drosophila* semaphorin I and II, *Tribolium* semaphorin I (087446) or
CC variola major (smallpox) virus semaphorin IV (087447) genes were used to
CC generate a series of peptides (R70370-R70416), which retain semaphorin
CC receptor binding activity. The semaphorin derived or semaphorin receptor
CC derived peptides are potent modulators of nerve cell growth, immune
CC responsiveness and viral pathogenesis. They can be used in diagnosis and
CC treatment of neurological disease and neuro-regeneration, immune
CC modulation and diagnosis and treatment of viral and oncological infection
CC and diseases.
SQ Sequence 724 AA:

Query Match	8.38;	Score 362;	DB 13;	Length 724;
Best Local Similarity	29.08;	Pred. No. 3.96e-22;		
Matches	84;	Conservative	68;	Mismatches 113; Indels 25; Gaps 19;

Db	265	vyffiretaaveylncgkavv-srlarvckkdvvgknlla-hmavtlykarlncsisgefp	322
Qy	208	lyfferefnndp-knpdearplnsvraolcrgdgoesslsvskwnrtfplkaalvcsdaatnk	266
Db	323	lyfneigevavgljpsdkarffat--ftctngljagsavcsfhineigaafgkfkgegsn	380
Oy	267	nfnrldovellfllpprscgmrtlrvlygvfnsnwnsnavcvstlsgldid--kv-fr---ts-	317
Db	361	sawlpvlinsvyebrpvtcyndtsnlbdvtlnlfrshpmdkavnh-eh-nmpvyktrdl	438
Oy	318	s--lkgvhssllpnrpgcldpdrptetfetoavadrhpvaorvempgklptlfhsk-y	374
Db	439	vfkrlvdkrlidllngeylyvuygtldtgytklykvuyrgesis-kllidfearneal	497
Oy	375	hykvaavahrmqas-hgefehnlylittbrcgtlhnkvvegeedehsnafnimeiorpprraa	433
Db	498	qmeisgtkrlsyigtldhrlkqldlmcnrltydn-cfcr-v-tdrucswd	544
Oy	434	qtmsladeerrklyvssowevsövpldlcew-yggcgagclmsrbycgwd	482

RESULT	15
ID	W57260 standard; Protein; 930 AA

DT 02-SEP-1998 (first entry)
DE Human semaphorin Y.
KW Human; semaphorin Y; nerve extension inhibitor; central nervous system;
KW peripheral nerve growth.
OS Homo sapiens.
PN MO9811216-A1.
PD 19-MAR-1998.
PF 09-SEP-1997; J03167.
PR 08-AUG-1997; JP-227220.
PR 11-SEP-1996; JP-263565.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-250958/22.
DR N-PSDB: V28915, V28916.
PT DNA encoding human and rat semaphorin Y - an inhibitor of nerve
PT extension
PS Claim 1: Page 65-70; 85pp; Japanese.
CC The present sequence represents human semaphorin Y, which inhibits
CC nerve extension. Semaphorin Y genes and proteins may be used to inhibit
CC peripheral nerve growth. Semaphorin Y antagonists can be used to
CC accelerate regrowth of the central nervous system.
CC Sequence 930 AA;

Query Match	8.18;	Score 352;	DB 32;	Length 930;
Best Local Similarity	28.48;	Pred. No. 3.40e-21;		
Matches 132; Conservative	112;	Mismatches 164;	Indels 56;	Gaps 44;

Db 120 ecnyyivtlvpwdsqclllaactgtsfspvcrsygltslqgeegelsgarcpcfatgtnva 179
 QY 93 DCEVYITLLERR-SEGGLACCTNRRHPSCNMLVNGTVAPLGE-MRGA--PFSADENSLV 148
 Db 180 lfaegsllysata-afgdsadavvyrslgppql-rs-a-kydskwlrphftvgslengdh 235
 QY 149 LFEDEYVISTRKQEVNGKIPR-FRRIRGESELYTSTQWONPEFIATVYHQDAVDK 207
 Db 236 vyffifreusvedarlqkvqfsrvaarvckirmgsprald-rhwstfklrlncs-vpds 293
 QY 208 IYFFREDNPDKNEADPLNRSVRAQLCRDQGGES-SLVSKNWTFKAMLVCSDATNK 266
 Db 294 tfy-f-dtqlglctprvnlhgrsa-lfyfttqtchslpagsavcafyldelergfegkikeq 355
 QY 267 NENRLOVEFL-LPDPSC-QWDRFVYVEFSNPWN-Y--SAVCYSLDDIDKVF--R-TS 317
 Db 351 rslgawtvpesdvpssprpscagvgaaalfssrclpddvllfkhaplldapvpvt 410
 QY 318 -SLWG-Y-H-S-S-LPMPRCKCLP-DQQRPIPTTFGVAAHNE-V-AQRY-EP-MPFLK 366
 Db 411 hqpllltstralltqvavdmgaphsn-ltwmfgsngdgtvklvlpdrgsggpepille 469

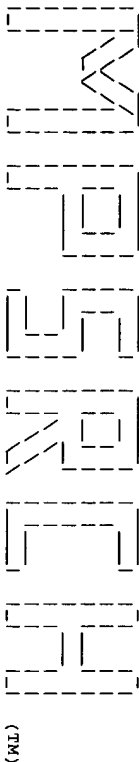
```

Oy 367 -TPLH--SKHYOKVAVRHQASGHEFNHLYLTTGRGTHHKVBERGEGEHSFAFNIME 423
Db 470 eIdaspaescskrtagrarrrlljleldtegrllfvafsgciyvjlsrcaariga--cqs 528
Oy 424 -IQP--R---RAA--A--IQMSDIAERKRLVSSQMEVSVPDLCEYVGCGCH-G 470
Db 529 clasgdyrcywhsrgygdvlgsgatdvdgagngesmehdcdg 572
Oy 471 CLMSDPICGMDQGR-CISITSSF-RSYLSINIAAE-PHKECPP 511

```

Search completed: Thu Jul 8 19:21:38 1999
Job time : 40 secs.

THIS PAGE BLANK (USPTO)



Release 3.1A John F. Collins, BioComputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 19:20:02 1999; Maspar time 26.93 seconds
Tabular output not generated. 864.549 Million cell updates/sec

Title: >US-09-041-236-2
Description: (26-606) from US09041236.pep (33 of 45)
Perfect Score: 4344
Sequence: 1 WKGHVGQDRVDFGQTEPHTV.....SYFEAGHWQLLPEDGIMAE 581

Scoring table:
PAM 150
Gap 11
Searched: 122810 segs, 40068593 residues
Post-processing: Minimum Match 08
Listing first 45 summaries

Database: p1r60
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 49.100; Variance 86.615; scale 0.567

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the change being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1850	42.6	653	2 T03102	semaphorin homolog A3	0.00e+00
2	737	17.0	748	2 I48744	semaphorin A - mouse	1.61e-134
3	725	16.7	751	2 I48748	semaphorin E - mouse	9.36e-132
4	716	16.5	749	2 G01856	semaphorin V - human	1.10e-129
5	688	15.8	753	2 G02173	semaphorin III family	2.96e-123
6	681	15.7	772	2 A49069	collapsin - chicken	1.19e-121
7	667	15.4	666	2 I58169	semaphorin III - mus	1.91e-118
8	669	15.4	772	2 I48747	semaphorin D - mouse	6.86e-119
9	661	15.2	771	2 D49423	semaphorin C - mouse	4.50e-117
10	618	14.2	782	2 I48746	semaphorin I - precu	2.87e-107
11	522	12.0	834	2 S66498	M-sema F protein prec	1.35e-85
12	480	11.0	760	2 I48745	semaphorin B - mouse	3.14e-76
13	470	10.8	403	2 E42521	hypothetical protein	5.19e-74
14	463	10.7	441	2 S29921	hypothetical protein	1.84e-72
15	461	10.6	730	2 JH0798	fasciclin IV precursor	5.09e-72
16	448	10.3	1074	2 JCS928	semaphorin I precursor	3.76e-69
17	389	9.0	711	2 A49423	semaphorin I - fruit	2.90e-56
18	367	8.4	656	2 B49423	semaphorin II precurs	1.89e-50
19	362	8.3	724	2 C49423	semaphorin II precurs	6.51e-36
20	293	6.7	295	2 J01775	SallpR protein - vacc	1.13e-16
21	196	4.5	1884	2 JC4975	plexin 2 precursor -	6.53e-12
22	170	3.9	1872	2 JC4976	plexin 3 precursor -	1.48e-11
23	168	3.9	1894	2 JC4980	plexin 1 precursor -	

24	171	3.9	1905	2 IS1553	plexin - African claw	4.32e-12
25	157	3.6	122	2 J01845	14R protein - variola	1.26e-09
26	157	3.6	122	2 H36852	A43R protein - variol	1.26e-09
27	155	3.6	142	2 J01776	SalpR protein - vacc	2.80e-09
28	120	2.8	775	2 E70320	polyrubonucleotide nu	1.37e-03
29	105	2.4	227	2 S09322	hypothetical protein	2.06e-01
30	105	2.4	235	1 Q0BEC9	HXLf4 protein - human	3.87e-01
31	103	2.4	406	2 B69064	conserved hypothetical	5.29e-01
32	102	2.3	275	2 A34866	T-cell surface proteol	1.33e+00
33	99	2.3	275	2 S08464	T-cell allonantigen RT	1.33e+00
34	99	2.3	460	2 S35772	translation elongatio	1.33e+00
35	99	2.3	563	2 S78224	virulence-associated	1.33e+00
36	99	2.3	591	2 S26565	virulence-associated	1.33e+00
37	99	2.3	591	2 S09498	virulence-associated	1.33e+00
38	99	2.3	591	2 S26664	virulence-associated	1.33e+00
39	99	2.3	593	2 S15215	virulence-associated	1.33e+00
40	99	2.3	597	2 A46050	thyroid/steroid recep	1.33e+00
41	98	2.3	641	2 S32017	flagellum-associated	1.80e+00
42	98	2.3	970	2 C57282	ankyrin-related prote	1.80e+00
43	98	2.3	979	2 B57282	ankyrin-related prote	1.33e+00
44	99	2.3	1375	2 JC5148	hepatocyte growth fac	1.33e+00
45	98	2.3	1786	2 A57282	ankyrin-related prote	1.80e+00

ALIGNMENTS

RESULT ENTRY	1	#type complete
TITLE	T03102	semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM	24-Mar-1999	#formal_name alcelaphine herpesvirus 1
DATE	24-Mar-1999	#sequence_revision 24-Mar-1999 #text_change
ACCESSIONS	T03102	
REFERENCE	Z14840	Ensser, A.; Pflanz, R.; Fleckenstein, B.
#authors	J. Virol. (1997) 71:6517-6525	
#title	Primary structure of the alcelaphine herpesvirus 1 genome.	
#accession	T03102	
##status	preliminary; translated from GB/EMBL/DBD	
##molecule_type	DNA	
##residues	1-653	##label ENS
##cross-references	EMBL:AF005370; NID:q2337967; PID:q2337970	
SUMMARY	#length 653 #molecular-weight 73645 #checksum 5501	
Query Match	42.68;	Score 1850; DB 2; Length 653;
Best Local Similarity	47.58;	Pred. No. 0.00e+00;
Matches	267; Conservative	95; Mismatches 188; Indels 12; Gaps 12;
Db	86 GQHRF-FGPOEPHTV-FHSLNSDYYVGGNNTIYLPFAHSSNASTALINTSTHTRH 144	
Qy	31 GQDRVFGQTEPHTV-FHEPGSSVWVGGRGKYLDFDEGKNASVRYNIGSTKSCLD 90	
Db	145 SSTCENFITLHNTDGLLACGTNSQKPSGW-LINNLTOTFLOPKLGLAFSPSSGNLV 203	
Qy	91 KRDCENYITLERRSGLLACGTNARHPSCMWLVNCTVVP-LGEMGVAPFSDENSLV 149	
Db	204 FPDNDYITINLKSLSGS-HKFRRIAGOVELYTSTAAHRPOFOVATVHKNESTDXT 262	
Qy	150 FEDDEVYSTIRK-OEYNGKIPRRIRGRSELYTSTVQNOQFIKATIVHDOQAYDXI 208	
Db	263 YFFEOGNSHDFKOPHTVPRYGOVCSDDGESSLSVYKMTFTFKARLACYDYDGRY 322	
Qy	209 YIFFRDNDKDKPEALNRSVAQLCRGQGGESSLSVSKMNTFTKALMLVCSDAATNKF 268	
Db	323 NELQDIFIQAPENSWEEITLYGLSPWNFSAVCFYTKVIDHYFKTSKLNYYHHRPT 382	
Qy	269 NRLQDVFLLPDPSGQRDRTRVYGVFSNPMNYSAYCVYSGDIDKDYRTSLGLYHSLPN 328	
Db	383 PRGQCKMKNHGHVPTTEFOVADRYPRVAPVYQKNNAMPITIOSKITYIKLIVRYEG- 441	
Qy	329 PRGKCLPQOQPIPTETFOVADRHPEVAQVPEMGLKPLRPHSKYHYQKAVVHRQASH 388	
Db	442 GYFWATIFLYTIKTGIHIVRYEDSNSTALNILEINFQRPAPQIONILLNTNLKLV 501	

QY	389	GETEH-VLYLTDDGTYTHKVVPEQEGHSFAFNIMEQPPRRRAAIDTMSLDEKRLIV	447
Db	502	NSEWESEVPLDLCYSGNDGFCSCFMSRDLPTIYNNTC-S-FK-QRVSYETGGPANTL	558
OY	448	SSQWEVSQVPLDCEYVGCGCGCLMSRDPYCGWDGRCISIVSESRVLSQINPAEPH	506
Db	559	SEMGCDHAPVYVHQSIPLLSSYLSICRAVSHADYFMTKGFTEKRCYVTHKNDCI	618
OY	507	KE-CPNKPPADKAPLOKQSLAPNSRYLSCPMESHATYSRHKENVEQSCGPGHSPNCI	565
Db	619	LIANSTATNGTHVCKNKEDS	640
OY	566	LFENTHMOOYGHYFCEAOGSS	587
RESULT	2		
ENTRY	148744	#type complete	
TITLE	semaphorin A - mouse		
ORGANISM	#format_name Mus musculus #common_name house mouse		
DATE	02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 04-Sep-1998		
ACCESSIONS	148744		
REFERENCE	148744		
#authors	Puschel, A.W.; Adams, R.H.; Betz, H.		
#journal	Neuron (1995) 14:941-948		
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension		
#cross-references	MUOID:95267431		
#accession	I48744		
#status	preliminary: translated from GB/EMBL/DBJ		
#molecule_type	mRNA		
#residues	1-748 #label RES		
#cross-references	EMBL:X85990; NID:9854323; PID:9854324		
GENETICS			
CLASSIFICATION	sema		
SUMMARY	#superfamily semaphorin		
	length 748 #molecular_weight 82894 #checksum 9017		
Query Match	17.0%; Score 737; DB 2; Length 748;		
Best Local Similarity	33.7%; Pred. No. 1,61e-134;		
Matches	168; Conservative 112; Mismatches 166; Indels 52; Gaps 36;		
Db	112	ECMNFVRLAHYNHTHLLACRTGAHFHTCALMRWATAGTTHASTGPEKLEDKGKTPDP	171
OY	93	DCENYITLERRSEG-LLACGTNRRHSC---WNLVNGT-V-V-P--LGEMGYAPFSP	142
Db	172	RHRPVSVALGELISGV-TADLMKRDITTRISLGQNSLTEDHDSRWLNKPFVAVFMT	230
OY	143	DENSLVLEFEGDEVSTIRKOEYNK-IPRRIRIRGESELYT-S-DIV-MONPOFIAT-I	197
Db	231	PESENPDDDKTYFFEFSEAVEAAPAMRMVSRYGQCRNDLGGQSL-VNKKTFPLKAR	289
OY	198	VHQDAIDDKTYFFREDNPDKNPE-APLVNSRYAQDLCRDQGGESLSYSKNITFLKAM	256
Db	290	LVCSVPEVEGDTHEDOIADYFLLS-SR-DRQTPLLAVFSTSSGVSFGASVICYSMNDYR	347
OY	257	LVCS-DAATNKNRNRIQDVFLLPDPGQQRDPYRVGVFSPNPN-V--SAYCYISLGIDID	311
Db	348	RAFLGFLPHKKGPRHONVSYOGKRPYTRPGMCSKSTFTGTSSKDPDDVYIQGRNHPLM	407
OY	312	KVFRTS-SLK-G-----YHSSLNRPBRPKCLPQD-QPIT-TEPE--R-HPEV	355
Db	408	YNPVLPMKG-R-PLFLQVAGYPTQTLAADRVAADSH-NDVLEICTDYGTVLYKVISVK	464
OY	356	AQRYEPMKPLTPLE-H-SKTHQKAAVHRMOSHEFTHTVLYTTDKTHKVE-P-410	
Db	465	GRPNSEGLLELEQVEDSAITSKQISSKROOLYVARSAAVAQIALHRTALGRCAE	524
OY	411	GEQHSFAFNIMEQPPRRRAAIDTMSLDEKRLIVSSQWEVSQVPLDCEYVGCGCG	470
Db	525	CCLARDPSCAMDSACTRFPQTPARFRERODIRNGDPSLCSGDSHSVLLKRYLVGES	584

Oy	471	CLMSRDPGCGMDGRCSTISYSSESSV--QSINPAPRHKECPKPKDKAPLQKVLAPRS	528
Dd	585	GSAPLECEPNSLOAHVOM	602
Oy	529	RYLLSCPMESRHATYSM	545
RESULT	3		
ENTRY	148748	#type complete	
TITLE	semaphorin E - mouse		
ORGANISM	#formal_name Mus musculus #common_name house mouse		
DATE	02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 04-Sep-1998		
ACCESSIONS	148748		
REFERENCE	148744		
#authors	Puschel, A.W.; Adams, R.H.; Betz, H.		
#journal	Neuron (1995) 14:941-948		
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension		
#cross-references	MUOJ:95267431		
#accession	148748		
#status	preliminary: translated from GB/EMBL/DBJ		
#molecule_type	mRNA		
#residues	1-751 #label RES		
#cross-references	EMBL:X85994; NID:g854331; PID:g854332		
GENETICS			
#gene	seme		
CLASSIFICATION	#superfamily semaphorin		
SUMMARY	#length 751 #molecular_weight 85259 #checksum 8961		
Query Match	16.7%; Score 725; DB 2; Length 751;		
Best Local Similarity	33.0%; Pred. No. 9,36e-132;		
Matches	148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;		
Dd	167	SEPNVNTVSMINIEELFSQMYI-DEMGIDDAFRSLTRMQLRTDQHSKMLSEPMFVD	225
Oy	139	PFSDEMSLVLEFEGDEYVSPTRKQENYNGK-IPFRRIKGESELYTS--DTV-MONPOFIK	194
Dd	226	AHVIPDGDNDADAVYEFKFERLTDNNRSRKQHSIARIICPNDDGQSRSL-VNKWTFEL	284
Oy	195	ATVYHQP-QAVDOKIYIFREDNDPKNPAPLAVSKVADLCGGDGGESSLSVKWNTFL	253
Dd	285	KARLVCSVTDEDEGETHFELEDFLL-ETDNP-RTTVLYGIGFTSSSVFKGSAYVYHL	342
Oy	254	KAMLVCS--DA-ATNKNFNRLOQVFLLPDSGGQWRTRYGVGSNPMWN-Y--SAVCYVSL	307
Dd	343	SDIOTVNGPRAHKEGNNHOLISYQGRIPPRRGCTPGGALFPNMRTTYKDPFDVDVYTR	402
Oy	308	GDIDCVFRTS-SLK-G-----YHSLPNNPKGCLPQD-QP-I-PIETF--QVADRRP	353
Dd	403	NHPLMYSNISPRIHRPLIVRIGTDYKTKIANDRVNAADG-RVHYVLCGTIDGTQOKVV	461
Oy	354	EVAQVRPEMGL-KTPI-F-H--SKHYQKAYAVHMAASGEFFHYLYLTDDGTIHKVYE	409
Dd	462	LPTNSSAGELLEELFEKFNHVPITTMETISKKQOLYVSSNEGVSQVSLARHCITGTAC	521
Oy	410	-PGQEHSPAFNIMEIPEFRRAAIDQMSLDAERKLYVSSQWEVSVQVPLDCEVYGGGC	468
Dd	522	ADCCLAADPCAMDGSHCSRFYTGKRRSRQDVRHGNPLTQCRGFNKLKAYNMAEIVQY	581
Oy	469	HGCLMSDPCMGMOGRCISIVSE--RSVLQSLINPAEPHKECP--NPKPKDKAPLQKVS	524
Dd	582	GVRNNSTFLECAPKSPQASIKWLLQDK	609
Oy	525	A-PNSRYLLSCPMESRHATYSM-RKEN	550
RESULT	4		
ENTRY	G01856	#type complete	
TITLE	semaphorin V - human		
ORGANISM	#formal_name Homo sapiens #common_name man		
DATE	21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 04-Sep-1998		

```

ACCESSIONS      G01856
REFERENCE        G08634
#authors        Sekido, Y.
#submission     submitted to the EMBL Data Library, June 1995
#accession      G01856
#status          Preliminary; translated from GB/EMBL/DDb
#molecule_type mRNA
#residues       1-749 #label SEK
#cross-references EMBL:U028369; NID:9974283; PID:9974284
CLASSIFICATION  #superfamily semaphorin
#length 749     #molecular-weight 83121 #checksum 2747
CUNYARY

```

Query Match	16.5%;	Score 716;	DB 2;	length 749;
Best Local Similarity	33.0%;	Pred. No. 1.10e-129;		
Matches	167;	Conservative 115;	Mismatches 170;	Indels 54;
				Gaps 41;

Db	112	ECMNFVKLHANYNTLHLLACGTGAFFPQCAVEVGHNAEERPLRLDGRITEDCKGSPYD	11
Oy	93	DCENITTLER-RBEGLLACCTNKRHSC-WNLV-N-G--TVPL--G--EM-RGAPFS	141
Db	172	PRHRAASVLGEELYSGVA-ADLMGRDFTTFRSLGQRPSLTERPHDSRWMLNEPKYKVEW	230
Oy	142	PDENSLVFEDEDEVYSTIRKOEYNGK-IPRRRIERGESELYT-S-DIV-MQNQFIKAT-	196
Db	231	IPESENPDODKITYEFFRETAVEAPALGRLSVSVGQICENDVGQSL-VNKNTEPIKA	289
Oy	197	IVHDDQAYDKITYIFFREDNDRKPE-APLNVSVAVOLCAGDQGGESSLSVSKMNTFLKA	255
Db	290	RLVCSVPGECDTHFDLOQVFLLS-SRDI-RPLLYAVFTSSSITFQGSVAVCSMNDY	347
Oy	256	MLVCS--DAATNKNFNRLQDFLLRDPDSGQMDRPIRVYGSNPMN-V--SAYCVYSIGDI	310
Db	348	RRATLGPANHEGSHQWVSVQGVPRPGMCSKTFGFSSKDPEDRDVIOFARNHPL	407
Oy	311	DK--V--F--RTSSL-K--GHSLPLMPRQCKLIPDO-QEIP-TEFF--OVAD--R-HPE	354
Db	408	MYSNVLPTGG-R-PLFLQVGANYTFTQIADRAVAADGH-YDVLFTIGDVTGLKYSVP	464
Oy	355	VAGVEEPGRPLKTEPLF-H--SKYHQKVAVHRMQASHGEFHVLXLTDTBGITHKVE-P	410
Db	465	KGSRPSAEGLLLEELNHFEDSAANTSMQISKRQLTVASRSVAVOALALRCAAHGRVCT	524
Oy	411	-GEBHSHAFENIMETIQPRRAAALQTSLSLAERKRLTVSSQMEVSQVPLDLCEVYGGCH	469
Db	525	ECCLLAPDYCAMDVACTREQPSAKRRFRRODVNCGPSTLCSDSSRPALLEHKEVGE	584
Oy	470	GCLMSRDPDYCCMDQGRCSISYSSRSVL-QSINPAPRHKECPRPKRDKAPL-QKYSIAP	526
Db	585	GSSAFLECEPRSLDARVWNTQORGAV	610
Oy	527	NSRYLSCPMESRHATYSWR-HKENV	551

RESULT	5	
ENTRY	G02173	#type complete
TITLE	semaphorin iii family homolog - human	
ORGANISM	#formal_name Homo sapiens #common_name man	
DATE	21-Dec-1996 #sequence_revision 06-Jun-1997	#text_change 04-Sep-1998
ACCESSIONS	G02173	
REFERENCE	G09275	
AUTHORS	Mayhew, S.	
#submission	submitted to the EMBL Data Library, October 1995	
#accession	G02173	
#status	preliminary; translated from GB/EMBL/DBEST	
##molecule_type	mRNA	
##residues	1-753	#label NAY
##cross-references	EMBL:U038276; NID:g1061350; PID:g1061351	
CLASSIFICATION	#superfamily semaphorin	
SUMMARY	#length 753	#molecular-weight 84941 #checksum 5681
Query Match	15.88;	Score 688; DB 2; Length 753;
Best Local Similarity	31.44;	Pred. NO. 2; 96e-123;

Matches 139; Conservative 121; Mismatches 140; Indels 42; Gaps 30;

Dy	171	PYDKLPDASALINEELIYAGYII--DFNGTDAIPRTGKOTARTDQYNSRLINDPFIH	229
Oy	139	PFSDEDSNLYFEDEDEYVSTRKEIYNGK-IPIRRRIKGESELYTS--DIV-HONPOFIK	194
Db	230	AELIPDASENDDKLYFFERKSAB-AEOSPAYVARIGRICLINDGGHCH-VNKMSTFLK	287
Oy	195	ATIHODQAYDCKIYTFEFREDNPKNDEAPLANSRVAQLCRGGDGGESSLSVKWMTFLK	254
Db	288	ARLYCSVPGEIGIETHDELDQVYV-QOTQDV-RNPIIYAVFTSSGVFSGSANVCYSMA	345
Oy	255	AMLVCS---DATAKRNENRLQDVFLLDPDSQWMDTVRYGVFSNP-WNY--SACVYSLG	308
Db	346	DIRVFNCPFAHKEGPNYQWMPFSGKMRPARGCPGFTFSPMSKSYDQVDEVEINFRS	405
Oy	309	DIDKVFRTS-SLK-G--YH-----SS-LPNRRPGKCLDQ-QP-I-PIETF--QVAD--R-	351
Db	406	HPLMYQAVYPLQRRPLVVRT-GABRYLTTTIAVDVDSADG-RKEVFLGTRGTQKVIY	463
Oy	352	HPEVAORPEPMG--PLKTPLFHSXYHOKYAVHMQSGHGFHFVLLTTRDGTIRKVE	409
Db	464	LPKDOEMEELMEVEVEFKPAVKYMTLSSKROOLYVASVGVTHLSLHRCQAVAAC	523
Oy	410	-PGDEHSPAFNINEIQOFFRPAALIQTMSLDAERKKLYVSQWMEVSOPPLDLCEYVGGC	468
Db	524	ADCCLARDPYCAMGOACSRYTASSKRRSRRODVRRHGNPIROCGRGFSNANKNAVESQY	583
Oy	469	HGCLMSRPGYCMQDGRGIS-IYSSER-SYLGSIINPAEPHKECF--NPKDPKAPLQKVSL	524

RESULT	6
ENTRY	A49069
TITLE	collapsin - chicken
ORGANISM	#formal_name Gallus gallus #common_name chicken
DATE	07-Apr-1994 #sequence_revision 07-Apr-1994 #text_changed 04-Sep-1998

ACCESSIONS	A49069
REFERENCE	A49069
#authors	Lu, Y., Raible, D., Raper, J.A.
#journal	Cell (1993) 75:217-227
#title	Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones.
#accession	A49069
#status	preliminary; not compared with conceptual translation
##molecule-type	mRNA
##residues	1-772 #label LUO
##cross-references	GB:U02528; NID:g410078; PID:g410079
CLASSIFICATION	#superfamily semaphorin
SUMMARY	#length 772 #molecular-weight 88867 #checksum 9712

[illegible]

```
QY 304 VYSLGIDDKVFRIS-SLK-G--YH--S--S-LPNRPCKCLPDC-QPI-PTETF--QV-A 349
Db 402 FARSHAPMYNVPFINS-RPIMIKTDVYDQTOIYVDRVAEDGQ-YDVMFIGDTGTVL 459
QY 350 -DR-HPEVAQVEBPMGLKTPLFHSK--YHYQKVAVHRMQASHGETFHVLYLTDRGTH 405
Db 460 KVASIRPKETHLEEVLEEMTVREPRVISAAMISTQOOLYIGSATGVQDLPLHRCDV 519
QY 406 KYVE-PEQEHSAFNFTME-IQPRRAAAIOTMSLDAERRLYVSSQWVSVPLDCEV 463
Db 520 YGKACACCCCLARDPYCAMDSSCSRYEPTAKRTRRDIRNGDPLTHCSDQHNDNPSGQ 579
QY 464 YGGCHCGLMSRDPYCCGDCGRCSISTYSSE-RSVL-QSINAPAEHKECPN-PRDKAP-- 518
Db 580 TLEEKITGVENSSTFLECSPKSQRAIVY-WQFOKONDHKEV 621
QY 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQCE 556

RESULT 7
ENTRY 158169 #type fragment
TITLE semaphorin III - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
04-Sep-1998

ACCESSIONS 158169
REFERENCE 158169
#authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;
#journal Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#title Neuron (1995) 14:949-959
#semaphorin III can function as a selective chemorepellent to
#pattern sensory projections in the spinal cord.
#cross-references MUID:95267432
#accession 158169
#status preliminary: translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-666 #label RES
#cross-references GB:L40484; NID:9703189; PID:9703190

GENETICS
#gene Semaphorin
#classification #superfamily semaphorin
SUMMARY #length 666 #checksum 9654

Query Match 15.4%: Score 667; DB 2; Length 666;
Best Local Similarity 30.1%: Pred. No. 1,91e-118;
Matches 150; Conservative 130; Mismatches 168; Indels 50; Gaps 39;

Db 60 RKGSPDPKLTASLLIDGELYSCTA-ANFMGRDFAIFRTLGHHPIRTEQDHSRWMLDP 118
QY 135 RGVAPFSPDESLVLFEGDEYVSTIRKOEYNGK-IPRRRIIRGESELYTS--DTV-WONP 190
Db 119 RFTSAHLIPESDNEDDKYFFFERENALDGEHSGKATHARIGQICKNDFGHRSL-VNKM 177
QY 191 OFIKATIVHQ-DQAYDRIKYFFREDNPDKNPEAPLVNSVAQLCRGOGESSLSYSKM 249
Db 178 TTFKALILCVSPRPNGLDTHFDELQVFLM-NSKDP-KNPITYGVFTTSSNIRKGSAYC 235
QY 250 NTFKAMLYCS--DA-ATNKNFNRLQDVFLLPDPGQWRDTRVYGVFSNPNV-Y--SAVC 303
Db 236 MYMSDVRVRLGPYAHARDGNVQWVPYGRGTPRPGTCSKTFGGGDSKTDLPDVIIT 295
QY 304 VYSLGIDDKV--F--RTSS-LK--GYHSSLPNRPCKCLPDC-QPI-PTETF--QV-A 349
Db 296 FARSHAPMYNVPFINS-RPIMIKTDVYDQTOIYVDRVAEDGQ-YDVMFIGDTGTVL 353
QY 350 -DR-HPEVAQVEBPMGLKTPLFHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTH 405
Db 354 KVASIRPKETHLEEVLEEMTVREPRVISAAMISTQOOLYIGSATGVQDLPLHRCDI 413
QY 406 KYVE-PEQEHSAFNFTME-IQPRRAAAIOTMSLDAERRLYVSSQWVSVPLDCEV 463
Db 414 YGKACACCCCLARDPYCAMDSSCSRYEPTAKRTRRDIRNGDPLTHCSDQHNDNHP 473
```

```
QY 464 YGGCHCGLMSRDPYCCGDCGRCSISTYSSE-RSVL-QSINAPAEHKECPN-PRDK--AP 518
Db 474 SLEERITGVENSSTFLECSPKSQRALVYWQFORNEDRKEIKMGDHIIRTOGLLRS 533
QY 519 -LQ-KVSLA-PNSRYLSC-PMESRHATY-SW-RHKEVNEQCEPG-HQ-SPNICLFIEN 570
Db 534 LQKDSGNYLCHAVEHGF 551
QY 571 LTAQYGHYFCEAOEGSY 588

RESULT 8
ENTRY 148747 #type complete
TITLE semaphorin D - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998

ACCESSIONS 148747
REFERENCE 148744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
#family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession 148747
#status preliminary: translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-772 #label RES
#cross-references EMBL:X85993; NID:9854329; PID:9854330

GENETICS
#gene Semaphorin
#classification #superfamily semaphorin
SUMMARY #length 772 #molecular_weight 88710 #checksum 1776

Query Match 15.4%: Score 669; DB 2; Length 772;
Best Local Similarity 30.2%: Pred. No. 6,66e-119;
Matches 170; Conservative 143; Mismatches 182; Indels 67; Gaps 51;

Db 58 HTFLDDE-ERSRLVYGAADHIFSFNLVINDFOKIWPVSY-TRDECKWAGKDIKECA 115
QY 43 HTVLFHBPSSSVYWGKRGKIVLFDPECKN-AS-VRTVINGSTKGSC-LDK---RDCE 95
Db 116 NFKIVLEAVYNOTHLXAGCTAFHPICITIEVGHHPEDNIRKLODSHFENGCRGSPDPKL 175
QY 96 NYTLLERRREG-LIACGTNARHPSCWNL-V--N--GTVPV--G--EM-RGVAPSPDE 144
Db 176 LTASLLIDGELYSCTA-ADMGRDFAIFRTLGHHPIRTEQDHSRWMLDPRFISAHILPE 234
QY 145 NSLVLFEGDEYVSTIRKOEYNGK-IPRRRIIRGESELYTS--DTV-WONFOIKATIVHQ 200
Db 235 SDNEEDDKYFFFERENALDGEHSGKATHARIGQICKNDFGHRSL-VNKTTFKARLIC 293
QY 201 -DQAYDRIKYFFREDNPDKNPEAPLVNSVAQLCRGOGESSLSYSKNTFLKMLVYC 259
Db 294 SVQCPNGLDTHFDELQVFLM-NSKDP-KNPITYGVFTTSSNIFKGSAYCMTSMQVRRV 351
QY 260 S--DA-ATNKNFNRLQDVFLLPDPGQWRDTRVYGVFSNPNV-Y--SAVCVYSLGIDKV 313
Db 352 FLGPYAHARDGNVQWVPYGRGTPRPGTCSKTFGGGDSKTDLPDVIITFGSHAPMYN 411
QY 314 FRIS-SLK-G--YH--S--S-LPNRPCKCLPDC-QPI-PTETF--QV-A-DR-HPEVAQ 357
Db 412 PVEPINS-RPIMIKTDVYDQTOIYVDRVAEDGQ-YDVMFIGDTGTVLVSVPEKTV 469
QY 358 RVEPMGLKTPLFHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTHKVE-PEQEH 414
Db 470 HDEEVLEEMTVREPRVISAAMISTQOOLYIGSATGVQDLPLHRCDIYGKACACCL 529
QY 415 HSFSAFNFTME-IQPRRAAAIOTMSLDAERRLYVSSQWVSVPLDCEVYGGGCHGCLM 473
Db 530 ARDPYCAMDSSCSRYEPTAKRTRRDIRNGDPLTHCSDLEHNDHNGHSLEERITGV 569
QY 474 SRDPYCGMDGRCISISTYSSE-RSVL-QSINAPAEHKECPN-PRDK--AP-LQ-KVSLA- 525
```

RESULT	ENTRY	ENTRY TITLE	ORGANISM	DATE	ACCSSIONS	REFERENCE	#authors	#journal	#title	#accession	#status	#molecule	#residues	#cros-references	GDB:SEMA1	#cros-references	#superfamily	#length	#molecular-weight	#checks	#sum
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	610																			
526	PNSRYLSC-PMESRHATYSR	546																			
9	10	148746	semaphorin C - mouse (fragment)																		
590	ENSSFECSKSORALYV-MQ	61																			

```

ORGANISM      #format_name Mus musculus #common_name house mouse
DATE          02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
              28-Feb-1997

ACCESSIONS    I48746
REFERENCE      I48746
#authors      Puschel, A.W.; Adams, R.H.; Betz, H.
#journal      Neuron (1995) 14:941-948
#title        Murine semaphorin D/collapsin is a member of a diverse gene
              family and creates domains inhibitory for axonal extension
#cross-references MIMD:55267431

#accession    I48746
#status       preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-782 #label RES
#cross-references EMBL:X83592; NID:9854327; PID:9854328

GENETICS
#gene         semC
SUMMARY       #length 782 #checksum 1571

Query Match      14.2% Score 618; DB 2; Length 782;
Best Local Similarity 33.7%; Pred. No. 2,876-107;
Matches 149; Conservative 96; Mismatches 148; Indels 49; Gaps 32;

Db
74 KRCONYIKILPLNSHLLTCGTAAPSICAVIHMSFTLADDEAGNVLTDEGKCHP 133
   |||:|:| | | | | | | | | | | | | | | | | | | | | | | | |
Qy 91 KRDEWITT-LLERSSEGLNCGTNAHPSC-W-N-----LV-NGTV-VPLGEMRTAPF 140
   |||:|:| | | | | | | | | | | | | | | | | | | | | | | | |

Db
134 DPNKSTALVVDGELLYGTVS-SFGQNDPAISRSQSSRPTKTESSLMLWDDPAFVASATS 192
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 141 SPDENSLVLFEGDVEVSTIRKQENKIPFRRIKGESELYTSDTV--MQNPQFIKATIV 198
   |||:|:| | | | | | | | | | | | | | | | | | | | | | | | |

Db
193 PESLGSPTGDDDKYIFFFSETGQFEFFENTT-VSRARVCKDGEGERVLYQ-QRMTSFL 250
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 199 HQDQ-A-Y--DCKLYFFREDNDP-KNPEAPLVYSRAQCRDQDQGESLSYSKWNTEFL 253
   |||:|:| | | | | | | | | | | | | | | | | | | | | | | | |

Db
251 KAOLLCSPDDGFEFNVLDVFTLNPNQD-DWRKTLISGVETSGOWHRTGTEGAIVCFM 309
   |||:|:| | | | | | | | | | | | | | | | | | | | | | | | |
Qy 254 KAMLVCSDAATNKFNFKLDQVFL-PPSCQMMDTRYGVFSNPMW--Y--SAVCYSL 307
   |||:|:| | | | | | | | | | | | | | | | | | | | | | | | |

Db
310 NDVQAFDGLYKKVNRFTQOYMETETHQVPRPGACITNSARFKINSLSQLDPRLVNF 369
   |||:|:| | | | | | | | | | | | | | | | | | | | | | | | |
Qy 308 GDIDKVF-----KISS-LKG-Y-HSS-LPMPRGKCLPD--QQPIPETQVADRPPEVA 356
   |||:|:| | | | | | | | | | | | | | | | | | | | | | | | |

Db
370 KDHFLLMDQVRSRLLLQPRARYQRAVAHVHVPGLHS-TYDVFELGTGDRLKAVALTSR 428
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 357 QRVBPM-GPLKTPLE--HSKYHQVAKVAHMQASHGETFVFLVLTDRGTHIKVPEEG 413
   |||:|:| | | | | | | | | | | | | | | | | | | | | | | | |

Db
429 VH---T-IEELQTPQOQPVQNLBDSHGGLLYASSHSIGVQVYPVANCISLYPT-CGDCLL 483
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 414 EHSFAFIMEIQPRRAAAIQTMSLDAERKRLVYSSQWEVSOYPLDCEVYGGGCGCLM 473
   |||:|:| | | | | | | | | | | | | | | | | | | | | | | | |

Db
484 ARDPYCAWTSAGACRLASTLYOPD 505
   :|:|:|:| | | | | | | | | | | | | | | | | | | | | | |
Qy 474 SRDPYCGMDGRC-I-SIYSSE 493
   :|:|:|:| | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
ENTRY   #type complete
TITLE   M-sema F protein precursor - mouse
ORGANISM #format_name Mus musculus #common_name house mouse
DATE    28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change
        10-Sep-1997
ACCESSIONS S66498
REFERENCE   S66498
#authors    Inagaki, S.; Furuyama, T.; Iwahashi, Y.
#journal    FEBS Lett. (1995) 370:269-272
#title      Identification of a member of mouse semaphorin family.
#cross-references MIMD:95385809
#accession  S66498
#status     preliminary
#molecule_type mRNA
#residues   1-834 #label INA
#cross-references EMBL:S79463; NID:91110598; PID:91110599

```


Db 302 EVIEKYNVLDIITKPLS 318
: : : : :
QY 347 QVADRHPVAVQRPVPMG 363

RESULT 14
ENTRY #type complete
TITLE hypothetical protein 15 - vaccinia virus
ORGANISM #formal_name vaccinia virus
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997

ACCESSIONS
#authors Amegeadzie, B.Y.
#submission submitted to the EMBL Data Library, January 1991
#accession S29921
#status preliminary
#molecule_type DNA
#residues 1-441 #label AME
#cross-references EMBL:X57318; NID:g62239; PID:g62234
SUMMARY #length 441 #molecular-weight 50185 #checksum 6034

Query Match 10.7% Score 463; DB 2; Length 441;
Best Local Similarity 33.9%; Pred. No. 1,84e-72;
Matches 87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;

Db 115 LVCGTNGNPKCMK-IDGSDPKHGRGYAPYQNSKVTIISHNGC-VLSDINISK-EG-I 170
: : : : :
QY 109 LACGTNARHPSQNLVNGTVPLGEMRGYAFSPDENSILVEGDEYVSTIRKQEVNGKI 168
: : : : :
Db 171 KWRREFDCCGYDLVADNVIPKDG-LRGAFYDKDGYT-VYIILFTDTIGSKR--I-VK 225
: : : : :
QY 169 PFRFRIRGE-S-ELYTSDVWQNPQFIKATIVHODQAYDKIYFREDNPKPEAPLN 226
: : : : :
Db 226 IYIOMCLNDGGPSSLSHRMSTFLKXVLEEC-DID-GRSY-R-Q-IHSRIKND-ND 279
: : : : :
QY 227 VSRVADLCGDDGESSLSVSKWNTFLKMLVCSDATNKNERRLDVFLDPDSCQMD 286
: : : : :
Db 280 TLIVYFEDSPYSKALCTYSMTIKOSFSTKLEGYTKOLPSPASGICLPAGKVVPHTF 339
: : : : :
QY 287 TNYGVGFSPMWYSAVCYSLDIDKVFRTSSLKGYHSSLPNRPQKCLPDQPIETEF 346
: : : : :
Db 340 EVIEKYNVLDIITKPLS 356
: : : : :
QY 347 QVADRHPVAVQRPVPMG 363

RESULT 15
ENTRY #type complete
TITLE fasciclin IV precursor - American bird grasshopper
ORGANISM #formal_name Schistocerca americana #common_name American bird grasshopper
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Sep-1997

ACCESSIONS
#authors JH0798
#journal Kolodkin, A.; Mathes, D.J.; O'Connor, T.P.; Patel, N.H.; Admon, A.; Bentley, D.; Goodman, C.S.
#title Neuron (1992) 9:831-845
#accession JH0798
#molecule_type mRNA
#residues 1-730 #label KOI
#cross-references GB:L00709; NID:g160844; PID:g160845
COMMENT #experimental_source embryo
This protein plays a role in growth cone guidance in the developing central nervous system.
glycoprotein; transmembrane protein

KEYWORDS
FEATURE
1-722 #domain signal sequence #status predicted #label SIG\
23-730 #product fasciclin IV #status predicted #label MAT\
23-627 #domain extracellular #status predicted #label EXT\
23-627

628-652 #domain transmembrane #status predicted #label TM\
653-730 #domain intracellular #status predicted #label INT\
44,71,163,267,360, #binding site carbohydrate (asn) (covalent) #status
539 predicted

SUMMARY #length 730 #molecular-weight 81214 #checksum 5881

Query Match 10.6% Score 461; DB 2; Length 730;
Best Local Similarity 28.0%; Pred. No. 5.09e-72;
Matches 143; Conservative 130; Mismatches 189; Indels 56; Gaps 46;

Db 32 QFGERVQRFGLNESHKDFKLEKDHNSLVGARNIYINISLRDLTEORTIMHSSG 91
: : : : :
QY 29 HVGODRVDFGOTEPHTVLFH--EPGSSSVWGGKGYLLFDFPQGN-ASVRYT-NI-G 82
: : : : :
Db 92 AHRELCTYKGSDEDCQONTIRYLAKIDDRVLICGTNAYKPLCRHYALKDGYVEKEYE 151
: : : : :
QY 83 STKSGC-L-DKR--DCENTITLRRSEG-LIACGTNARHPSQNLVNGT-VVPLG-E 133
: : : : :
Db 152 GGLCPFPDPDNSTAIYSEGQLYSATV-ADFGSDP-LI-YRG-P-LRTERSDL-K-QLN 204
: : : : :
QY 134 MRYAPFSPDENSILVEGDEYVSTIRKQEVNGKI-PFRRIIGSEELVTSOTVMQNPQFI 193
: : : : :
Db 205 APNEYN-TMEYNDLTFEFPRETAVEYINGKAIT-SRYARVCKHDKGPHQFG-DRMTSF 261
: : : : :
QY 194 KATIVHODQAYDKIYFREDNPD-KNEAPLNVSRVAQLCRGQGGSSLSVSKWNTF 252
: : : : :
Db 262 LKSLNCSVPGDYPPYFNEIQTSDIISNVGO-VKLIYGFVFTTPVNSIGSAVCAFS 320
: : : : :
QY 253 LKAMLVCSDAATNK-NFNRLQDF-LLPDP-SGQMRDTRVGVFSPNPN-Y-SAVCYAS 306
: : : : :
Db 321 MKSILESPDGPKEQETMNSNMLAVPSLKVPPEPRGQCVNSRTLPDVSNVFKSHITLMD 380
: : : : :
QY 307 L-G--D-IDKVFRT--SLKG-YHS--SL--PMPRGKCLPDQPIETEFQVADNHPVYA 356
: : : : :
Db 381 EAVPAFFT-RPILIRISLQYRTKTAVDQVTPGKAYDLVFICTDGGKVIKALNSAF 439
: : : : :
QY 357 QVPEPMGPLKTPLEFH-S-KYHOKYAV-HRMQASHGETFHVLYLTDTGKTIHKVVEPG-- 411
: : : : :
Db 440 DSSDVIDSVIEELQVLPFGVPYKMLVYRMGDGDSKLVVSDDLIAIKLRGSDKIT 499
: : : : :
QY 412 EGHSHFAFNIME-IQFRRAAIQT---SLDAERRKLYVSSQWESVQVPLDCEVYG-G 466
: : : : :
Db 500 NCRECVSLQDPYCAMDNELKCTAGVSPDMS 530
: : : : :
QY 467 GCHGLMSRDPYCGWDQ-G-RCISITSSERS 495

Search completed: Thu Jul 8 19:20:38 1999
Job time : 36 secs.

THIS PAGE BLANK (USPTO)

Db	280	TLVYEFSPSKALCYKSNMTIKQSTKSLGSGYTKQLSPSPSGICLPKGVAPHTF	339
Oy	287	TRVGVSENPWNYSAVCYSLGCDIDKVFRTSLGLGYHSSLPNPRGCKLPDQOPPIPTF	346
Db	340	EVIEKYNVLDDIIRPLS	356
Oy	347	QVADHPPEVAQREPMG	363
RESULT	3		
ID	SEX_HUMAN	STANDARD:	PRT: 1871 AA.
AC	P51805.		
DT	01-OCT-1996 (REL. 34, CREATED)		
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)		
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)		
DE	TRANSMEMBRANE PROTEIN SEX PRECURSOR.		
GN	SEX.		
OS	HOMO SAPIENS (HUMAN).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;		
OC	PRIMATES; CARNIVORINIA; HOMINIDAE; HOMO.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=FETAL BRAIN, SKELETAL MUSCLE, AND EMBRYO;		
RC	MEDLINE: 96149362.		
RA	MAESTRINI E., TAMAGNONE L., LONGATI P., CREMONA O., GULISANO M.,		
RA	BIONE S., TANANINI F., NEEL B.G., TONIOLO D., CONOGLIO P.M.:		
RT	"A family of transmembrane proteins with homology to the		
RL	MET-hepatocyte growth factor receptor."		
CC	PROC. NATL. ACAD. SCI. U.S.A. 93:674-678(1996).		
CC	- FUNCTION: PUTATIVE RECEPTOR INVOLVED IN THE DEVELOPMENT OF NEURAL		
CC	AND EPITHELIAL TISSUES.		
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).		
CC	- TISSUE SPECIFICITY: PREDOMINANT IN NEURAL TISSUE DURING		
CC	DEVELOPMENT (BY SIMILARITY); WIDELY EXPRESSED IN ADULT TISSUE.		
CC	- SIMILARITY: WITH THE EXTRACELLULAR DOMAIN OF THE MET/RON/HGF		
CC	RECEPTORS AND THE CYTOPLASMIC DOMAIN OF SEP, NOV AND OCT.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation		
CC	at the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X87852; E183447; -.		
DR	MM; 300022; -.		
KW	RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.		
FT	SIGNAL	1	19
FT	CHAIN	20	1871
FT	DOMAIN	1221	1220
FT	TRANSNEM	1241	1241
FT	DOMAIN	1242	1871
FT	CARBOHYD	59	59
FT	CARBOHYD	548	548
FT	CARBOHYD	637	637
FT	CARBOHYD	738	738
FT	CARBOHYD	746	746
FT	CARBOHYD	1009	1009
FT	CARBOHYD	1036	1036
FT	CARBOHYD	1073	1073
FT	CARBOHYD	1115	1115
FT	CARBOHYD	1162	1162
SO	SEQUENCE	1871 AA;	207661 MW; 9411046A CRC32;
Db	427	VVFICGTSNGSLKKYRVDFQDAHL-YETVPVYD--GSPILDLFLSPDHRHIIYLSEKOV	483
Oy	394	VLYLTDTGTLTHKVFPEQDEHSFAFNMELOPFRRAAIIOTMSLDLDERKRLVYSSOMEV	453

[illegible]

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 95266321.
RA GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RT MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.:
"The DNA sequence of human herpesvirus-6: structure, coding content,
and genome evolution."
RL Virology 209:29-51(1995).
CC -I- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
REPLICATION.
CC -I- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -I- SIMILARITY: TO OTHER HERPESVIRUSES' DNA-BINDING PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X92436; G1044871; -
DR EMBL: X83413; G854020; -
DR PFM: P00747; viral_DNA_bp; 1
DR DNA-BINDING: DNA REPLICATION; ZINC-FINGER; NUCLEAR PROTEIN.
FT ZN_FING 459 475
FT C4_TYPE.
SQ SEQUENCE 1132 AA; 127762 MW; DB699A02 CRC32;

Query Match 2.3%; Score 98; DB 1; Length 1132;
Best Local Similarity 28.2%; Pred. No. 4.38e-01;
Matches 24; Conservative 18; Mismatches 37; Indels 6; Gaps 6;

Db 435 WNLNRKSYNAGNAHNEI-VNHLVNCNANCFRCGKC-C-OSGIGTAMVRGTRLPAP 491
OY 297 WNTSANCYVSLGDIDKVFRTSLKGYHSSLPNRPCKCLPDQPIETETQVADRPEVA 356
DB 492 KNYK-KEPLVMSKF-SRY-YAEVDI 513
OY 357 QREVPKPLKTLPLFHSKYHYQKAV 381

RESULT 14
ID HA21_HUMAN STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DE 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84168117.
RA AUFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPEL D.,
RT STROMINGER J.L.:
"Isotypic and allotypic variation of human class II
RT histocompatibility antigen alpha-chain genes."
RL NATURE 308:327-333(1984).
RN [2]
RP SEQUENCE OF 40-255 FROM N.A.
RX MEDLINE: 83065171.
RA AUFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.:
RT "cDNA clone for the heavy chain of the human B cell alloantigen DC1:
RT strong sequence homology to the HLA-DR heavy chain."
RL PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
RN [3]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

entites requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: J00199; -; NOT_ANNOTATED_CDS.
DR PIR: A02212; HLHDC.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFM: PF00047; 1; 1.
DR PFM: PF00993; MHC_II_alpha; 1.
DR HSSP: P01910; 11AK.
KM MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 23
FT CHAIN 24 255
FT
FT DOMAIN 24 110 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 111 204 EXTRACELLULAR ALPHA-1,
FT DOMAIN 205 217 EXTRACELLULAR ALPHA-2,
FT TRANSMEM 218 240 CONNECTING PEPTIDE.
FT DOMAIN 241 255 CYTOPLASMIC TAIL.
FT DISULFID 133 189 BY SIMILARITY.
FT CARBOHYD 104 104 POTENTIAL.
FT CARBOHYD 144 144 POTENTIAL.
SQ SEQUENCE 255 AA; 28105 MW; 5C46DD59 CRC32;

Query Match 2.2%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1.20e-00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;

Db 25 DIVADSVADLG-VNLQSGSPSGYSHEFDGDEFLYVDKERETWQLFLFRFR 78
OY 122 NLVNGTVPLGEMRGVAPSPSPDENSIVLEFGDEV-YSTIRKQELNGKIRFRFR 175

RESULT 15
ID BSR_ECOLI STANDARD; PRT; 329 AA.
AC P2551;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOSE OPERON REPRESSOR.
DE BSR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6 AND 250-256.
RC STRAIN-K12.
RX MEDLINE: 93278299.
RA MAUZY C.A., HERMODOSON M.A.:
RT "Structural and functional analyses of the repressor, Rbsr, of the
RT ribose operon of Escherichia coli."
RL PROTEIN SCI. 1:831-842(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RC MEDLINE: 93315143.
RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.:
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication."
RL GENOMICS 16:551-561(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12.
RA ITIDA A., TESHIBA S., MIZOBUCHI K.:
RT SUBMITTED (XXA-1993) TO EMBL/GENEBANK/DBJ DATA BANKS.
RN [4]
RP SIMILARITY TO RIBOSE-BINDING PROTEINS.
RX MAUZY C.A., HERMODOSON M.A.:
RT "Structural homology between rbs repressor and ribose binding protein
RT implies functional similarity."
RL PROTEIN SCI. 1:843-849(1992).
CC -I- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RSDACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING

```
CC THE RBS OPERON TRANSCRIPTIONAL START SITE. THE AFFINITY FOR THE
CC RBS OPERATOR IS REDUCED BY ADDITION OF RIBOSE, CONSISTENT WITH
CC RIBOSE BEING THE INDUCER OF THE OPERON.
CC -1- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13169; G147517; -.
CC EMBL; L10328; G290603; ALT_INIT.
CC EMBL; AE000452; G1790194; -.
CC EMBL; D10466; G471110; -.
CC PTR: A41828; A41828.
CC ECGENE; EG10819; RBSR.
CC PROSITE; PS00356; HTH_LACI_FAMILY; 1.
CC PFAM; PF00356; lacI; 1.
CC PFAM; PF00532; Peripla_BP_like; 1.
CC HSSP; P15039; IPRV.
CC KW TRANSCRIPTION REGULATION; REPRESSOR; DNA-BINDING.
CC INIT_MET 0
CC FT DNA_BIND 3 22 H-T-H MOTIF (BY SIMILARITY).
CC FT CONFLICT 11 12 GV -> L (IN REF. 1).
CC SO SEQUENCE 329 AA; 36480 MW; 424E72A6 CRC32;
CC
CC Query Match 2.2%; Score 94; DB 1; Length 329;
CC Best Local Similarity 34.6%; Pred. No. 1.67e+00;
CC Matches 18; Conservative 14; Mismatches 14; Indels 6; Gaps 6;
CC
CC Db 109 TLMOKRVVDGLLLC-TETHOPS-REIMQRYPTVP-TVMMDNAPFDGSD-LI 156
CC |:::|::| |:::|::| |::|::| |::|::| |::|::| |::|::|
CC Qy 99 TLERRSEGLLA-CGTNRHPSCHWLVNG-TVVPDGENMGYAFSPDENSELY 148
```

Search completed: Thu Jul 8 19:18:33 1999
Job time : 21 secs.

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:18:51 1999; MasPar time 37.97 Seconds

Tabular output not generated. 835.158 Million cell updates/sec

Title: >US-09-041-236-2

Description: (26-606) from US09041236.pep (33 of 45)

Sequence: 1 WKGHVGQDRVDFGQTEPHTV.....SYFREAHQWQLPEDGIMAE 581

Scoring table: PAM 150

Gap 11

Searched: 179066 segs. 54579741 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

sptemb19
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 48.661; Variance 75.239; scale 0.647

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	4344	100.0	666	4	075326	0.00e+00
2	2273	52.3	393	11	088371	0.00e+00
3	1850	42.6	653	14	064906	0.00e+00
4	737	17.0	748	11	062177	5.72e-154
5	725	16.7	751	11	062181	8.15e-151
6	716	16.5	749	4	013214	1.88e-148
7	716	16.5	750	4	093018	1.88e-148
8	710	16.3	751	4	099985	7.05e-147
9	705	16.2	751	13	042236	1.44e-145
10	688	15.8	753	4	013372	4.09e-141
11	682	15.7	754	11	088633	1.52e-139
12	681	15.7	772	13	090607	2.77e-139
13	682	15.7	785	11	088632	1.52e-139
14	681	15.7	785	4	012725	2.77e-139
15	679	15.6	785	4	015704	9.25e-139
16	669	15.4	772	11	062180	3.79e-136
17	668	15.4	772	11	063548	6.97e-136
18	661	15.2	771	4	014563	4.65e-134
19	618	14.2	782	11	062179	7.20e-123
20	614	14.1	775	11	P70275	7.85e-122

21	599	13.8	861	11	009126	SEMAPHORIN J (SEMAPHOR	6.05e-118
22	594	13.7	761	13	090663	COLLAPSID-2.	1.19e-116
23	592	13.6	775	4	015041	KIA00331.	3.92e-116
24	572	13.2	785	13	042237	COLLAPSID 5.	5.71e-111
25	568	13.1	294	13	090664	COLLAPSID-3 (FRAGMENT)	6.14e-110
26	544	12.5	862	4	092854	SEMAPHORIN-3 (FRAGMENT)	9.11e-104
27	522	12.0	834	11	064151	SEMAPHORIN I (M-SEMA F	3.93e-98
28	511	11.8	299	13	090666	COLLAPSID-5 (FRAGMENT)	2.53e-95
29	480	11.0	760	11	062178	SEMAPHORIN-B PRECURSOR	1.93e-87
30	473	10.9	295	13	090665	COLLAPSID-4 (FRAGMENT)	1.14e-85
31	461	10.6	730	5	026473	FASCICLIN IV	1.23e-82
32	448	10.3	1074	4	013591	SEMAPHORIN F HOMOLOG.	2.31e-79
33	441	10.2	1077	11	062217	SEMAPHORIN F PRECURSOR	1.32e-77
34	416	9.6	1093	11	060519	SEMAPHORIN G PRECURSOR	2.37e-71
35	390	9.0	494	4	060408	SEMAPHORIN F (FRAGMENT	6.71e-65
36	392	9.0	888	11	035464	SEMAPHORIN VIA.	2.15e-65
37	378	8.7	712	5	026972	SEMAPHORIN-I PRECURSOR	6.11e-62
38	367	8.4	562	5	017330	CESEMA.	3.07e-59
39	367	8.4	771	5	024322	SEMAPHORIN-I.	3.07e-59
40	362	8.3	706	5	024323	SEMAPHORIN-II.	5.13e-58
41	341	7.8	284	11	054948	SEMAPHORIN IV HOMOLOG	6.64e-53
42	316	7.3	886	11	054951	SEMAPHORIN N (SEMAPHOR	6.95e-47
43	316	7.3	887	11	070141	SEMAPHORIN Z	6.95e-47
44	283	6.5	770	5	044253	SIMILAR TO SEMAPHORIN-	4.57e-39
45	204	4.7	1963	4	075051	KIA00463 PROTEIN (FRAG	3.56e-21

ALIGNMENTS

RESULT	ID	1	PRELIMINARY:	PRT:	666 AA.
AC	075326				
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)				
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)				
DE	SEMAPHORIN L.				
GN	SEMAPHORIN L.				
OS	HOMO SAPIENS (HUMAN).				
CC	EUAROTIA; METAFOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;				
OC	CARARHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98399619.				
RA	LANG E C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;				
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA				
RL	viruses".				
DR	GENOMICS 51:340-350(1998).				
DR	EMBL; AF030698; G3523115; -.				
SO	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;				

Query Match	Best local similarity	100.0%	Score 4344;	DB 4;	Length 666;
Matches	581;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;		
Db	58	WKGHVGQDRVDFGQTEPHTVLFHEPGSSVWVGKGYLDFPEGKNASRYTNIGSTK	117		
Oy	26	WKGHVGQDRVDFGQTEPHTVLFHEPGSSVWVGKGYLDFPEGKNASRYTNIGSTK	85		
Db	118	GSCLDKRCCEVITLLERSSGLACGTNARHPCWNLVNTVPLGEMRGYAPSPDEN	177		
Oy	86	GSCLDKRCCEVITLLERSSGLACGTNARHPCWNLVNTVPLGEMRGYAPSPDEN	145		
Db	178	SLVFESEGEVYSTRKQYKIRPRIRIRESSELYSDTYMNPQFIKATVHQDAYD	237		
Oy	146	SLVFESEGEVYSTRKQYKIRPRIRIRESSELYSDTYMNPQFIKATVHQDAYD	205		
Db	238	DKIYFFREDNDPKNEAPLVNSVAQLCRDQGESLSYSKNTFLKALVCSDATN	297		
Oy	206	DKIYFFREDNDPKNEAPLVNSVAQLCRDQGESLSYSKNTFLKALVCSDATN	265		
Db	298	KNFNRLQDVFLLPDPSGQMDRTRYGVFSNPNWNSAVCVYSLGIDIKVFTSLKGYHSS	357		
Oy	266	KNFNRLQDVFLLPDPSGQMDRTRYGVFSNPNWNSAVCVYSLGIDIKVFTSLKGYHSS	325		

Db	LEPRPGRKLPPOORLPETTFQVADNHREVAQVVERMGRKTPFLFSKYNHVKVAANRRHQ	417
Oy	LNPRGRKLPPOORLPETTFQVADNHREVAQVVERMGRKTPFLFSKYNHVKVAANRRHQ	385
Db	ASHGETFHLVLTDTGRTLHKVVERPEQDEHSEAFNIMEIQPFRAAAIQTMSIDAERRKL	477
Oy	ASHGETFHLVLTDTGRTLHKVVERPEQDEHSEAFNIMEIQPFRAAAIQTMSIDAERRKL	445
Db	YSSQMEVSVQVLDLCEVYGGGCHGCLMSRDPCGMDOGRCTISSESSVLOSINPAER	537
Oy	YSSQMEVSVQVLDLCEVYGGGCHGCLMSRDPCGMDOGRCTISSESSVLOSINPAER	505
Db	HKECNPGRKDKAPLQKVLAPNSRYVLTSCPMESRNATYSMRKHNENQOCEPQHSPNCI	597
Oy	HKECNPGRKDKAPLQKVLAPNSRYVLTSCPMESRNATYSMRKHNENQOCEPQHSPNCI	565
Db	LFLENLTAQVGHYGFCEADQEGSTFRRAQGMOLLPRGCIINAE	638
Oy	LFLENLTAQVGHYGFCEADQEGSTFRRAQGMOLLPRGCIINAE	606

RESULT	2	PRELIMINARY:	PRT:	393 AA.
ID	088371			
AC	088371:			
DT	01-NOV-1998 (TREMBLREL, 08, CREATED)			
DT	01-NOV-1998 (TREMBLREL, 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL, 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L (FRAGMENT).			
GN	SEMA.L.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUFAROTA; METAQOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 98389619.			
RA	LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;			
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA			
RT	viruses."			
RL	GENOMICS 51:340-350(1998).			
DR	EMBL: AF030689; G3523117; .			
FT	NON_TER 393			
SO	SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;			
	Query Match	52.3%:	Score 2273:	DB 11; Length 393;
	Best Local Similarity 91.9%:		Pred. No. 0.00e+00:	
	Matches 307; Conservative 14; Mismatches 12; Indels 1; Gaps 1			
Db	60 GODHVSQPEPHVTLFHEPGSFVWVGGRKVVHFNPECKGNASVRTVNGSTKSGCOD 119			
Oy	31 GODRVDFQTEPHVTLFHEPGSSVWVGGRKVVTLFDEPECKNNSVRTVNGSTKSGCSD 90			
Db	120 KODCNGTITLLERNGNGLVCGTNARAKSCNNLVNDSVYNSLGEKKGTAPSPDENSLVL 179			
Oy	91 KRDCENYITLLERSEGLACGTNARHPSCNNLVNGTVV-PLGEMRGYAPFSPDENSLVL 149			
Db	180 FEEDGVSYTIKROENKGIIPRRFRIRGSESLYTSDTWQNPQFIKATIVHDDQAYDDKIY 239			
Oy	150 FEEDGVSYTIKROENKGIIPRRFRIRGSESLYTSDTWQNPQFIKATIVHDDQAYDDKIY 209			
Db	240 YFFREDNDKNPDEAPLVNYSRYAOLRCGDOGESSLSYKNTFLKAMLVCSDAATNRNFN 299			
Oy	210 YFFREDNDKNPDEAPLVNYSRYAOLRCGDOGESSLSYKNTFLKAMLVCSDAATNRNFN 269			
Db	300 RLQDVFLLPDSGQMRDTRVYGVFSNPWNYSAVCVYSLGIDIDRVFTSSLKGVMGLSNP 359			
Oy	270 RLQDVFLLPDSGQMRDTRVYGVFSNPWNYSAVCVYSLGIDIDRVFTSSLKGVMGLSNP 329			
Db	360 RPKGKLPKKQPIPTETFEVADSHPEVAORVPMG 393			
Oy	330 RPKGKLPDQQPIPTETFEVADRHPEVAORVPMG 363			

ID	Accession	Source	Size	Score	DB	Length
3	064906	PRELIMINARY	653 AA			
AC	064906					
DT	01-NOV-1996	(TREMBLREL, 01, CREATED)				
DT	01-NOV-1996	(TREMBLREL, 01, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998	(TREMBLREL, 08, LAST ANNOTATION UPDATE)				
DE	SIMILAR TO GENBANK ACCESSION NUMBER L26081.					
OS	ALCEPHINE HERPESVIRUS 1.					
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;					
OC	GAMMAHERPESVIRINAE.					
RN	1					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-C500;					
RC	MEDLINE: 97201573.					
RA	ENSER A., FLECKENSTEIN B.;					
RT	"Alcephine herpesvirus type 1 has a semaphorin-like gene.";					
RL	J. GEN. VIROL. 76:1063-1067(1995).					
RN	2					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-C500;					
RC	MEDLINE: 97404659.					
RA	ENSER A., PELANZ R., FLECKENSTEIN B.;					
RT	"Primary structure of the alcephine herpesvirus 1 genome.";					
RL	J. VIROL. 71:6517-6525(1997).					
RN	3					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-C500;					
RC	ENSER A., PELANZ R., FLECKENSTEIN B.;					
RA	SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.					
DR	EMBL, U18243; G1000717; -					
DR	EMBL, AF005370; G2337970; -					
SO	SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;					
Query Match	42.6%;	Score 1850;	DB 14;	Length 653;		
Best Local Similarity	47.5%;	Pred. NO. 0.00e+00;				
Matches 267;	Conservative 95;	Mismatches 188;	Indels 12;	Gaps 122;		
Db	86	GOHRR-FGPOEPHYVFLSLNSSDVYVCGNTTYLFPFAHSSNASTALINTSTHNTHL	144			
Qy	31	GODRDFEQTEPHYLFHEPSSSSVWVGKRYLLFDFPEKKANSVTVNIGSTKGSCLD	90			
Db	145	SSTCENFTLLHNQTDGLACGTSQKPSCW-LINNITTOFLAGKLGAPSPSSGNLYL	203			
Qy	91	KRCDCNYITLLERSEGLACGTNARHPSCNVLNGVTVP-LGEMRGVAPSPDENSIVL	149			
Db	204	FDQNTYTSITNLKYLSSS-HKFRRIAGQVLLYTSPTAMHPQVQVQATAHKNSYDCKI	262			
Qy	150	FEGDEYVSTIRK-OEYNKRIEPRFRIRSESLYSTDVYMONPOEIKATIVHQDAYDCKI	208			
Db	263	YFFQENSHSPFKOPPHVPRVGVGSSDQGESLSVYKTTFLTKRLKACVDVDTGRIT	322			
Qy	209	YFFEDNPDKNPPLAVNSKVAQLCRDQGESLSVSKNTFLKMLKLYCSDPATIKNF	268			
Db	333	NELQDIFIQAPENSMEETLLYGLFSLPWFSAVCEFTVKDIDVFTSKLNYHNR.LPT	382			
Qy	269	NRLQVFLPDPSCGMDRTVRYGVFSNPMNYSAVCVYSLGIDIDKVFRTSLKGYHSLPN	328			
Db	363	PRPGCGMNHQHPVETQVADRPREVADPRYQKNNMFP.LIQSKYITTKLYRVEXG-	441			
Qy	329	PRPGCLDDQDPIPETFTQVADRHEVAQRAQEPKPLKTFHSHKRYQKVAHYRMQASH	388			
Db	442	GVFATIEYLLTIKTIHYYRYEDSNSTALINLEINPEQKRPVIOINLLDNTNLELY	501			
Qy	369	GETPH-VYLITTDKGTIKHYVEPQDEHSFANFMELQPPFRRAAIDTMSLDAARRKLY	447			
Db	502	NSEWSEVSEPLDLCSSVYGNDSCSEFMSHPDLCTWYNNTC-S-EK-QRYSVETGGRANRL	558			
Qy	448	SSQWEVSQVPLDCEVYGGCGHCLMSDPYCGWDQGRKISITYSEERSVLOSINPAPH-	506			
Db	559	SEMCGDHAIPVYVHQSIP.LLSNYSYLSICRAVSNHADIYFWTKDGTFRKCHVTKHKDCI	618			
Qy	507	KE-CENPAPKAPLQKYSLANSRYSYLLSCPESSHATYVSMKHEKNVQSCPEGHQSPNCI	565			

OY 410 -GEGEHSAFNAIMETIOPERRAAAIQTMSLDERRKLYSSQMEVSQVPLDCEYVGGGC 468

Db 522 ADCCLARDPYCAMDGHSCRFYPTGKRSRRODVNRHGNPLTQCRGNFLAYRNAAEIVQY 581

OY 469 HCCGLMSRDPYCGMDQRCISITSSSE--RSVLOSINPAEPHKECP--NPKRPAKLOKVSU 524

Db 582 GVRNNSTFLECAPKSPQASIKWLLQDK 609

OY 525 A-PNSRYLSCPMESRHATYSWR-HKEN 550

RESULT 6

ID Q13214 PRELIMINARY: PRT: 749 AA.

AC Q13214;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE SEMAPHORIN V.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE; 96210603.

RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H., ALBANESE J.P., LEE C.C., LERMAN M.I., MINNA J.D., "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung cancer deletion region and demonstrate distinct expression patterns."

RT PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).

RL EMBL; U28369; G974284.

DR PFAM; PF00047; 19; 1.

SO SEQUENCE 749 AA; 83121 MW; 6138C5D5 CRC32;

Query Match 16.5%; Score 716; DB 4; Length 749;
Best Local Similarity 33.0%; Pred. No. 1,886-148;
Matches 167; Conservative 115; Mismatches 170; Indels 54; Gaps 41;

Db 112 EGMNFVKLLHAYNRTHLACGTGAHPHCAPVEVGHRAEPEVRLRDPGRIDGKSPYD 171

OY 93 DCENITTLER-RSEGLLACGTINARHPS-C-WNLV-N-G--TVVPL--G--EM-RGIAPFS 141

Db 172 PRHRAASVLYGELYSGVA-ADLMGRDFTIFRSLGQRPSLRTPEHDSRWLNPEKVFYEW 230

OY 142 PDENSLVLEGEDEVYSTIRKQEVNKG-IPRFRIRGESELYT-S-DTV-MQNPQFIKAT- 196

Db 231 IPESNPDDDKTYFFRFRTEAVALRSLYSRNGQICRNDVGGORSL-VNKTTFELKA 289

OY 197 IVHODQAYDDKTYFFREDNPDKNPE-APLNVSRVAQICRDOGESSLSYSKWNTEFLKA 255

Db 290 RLVCSPVGEQDTHFDOLDFLLS-SRDH-RTPLLYAVFSTSSIFQCSAVCYVSMNDV 347

OY 256 MLVCS--DAATKNKRNRLQDVFLLRDPSCGMROTIRYGVFSNPMN-Y--SAVCYISLGD 310

Db 348 RRAFLGPRANKEGPMHGVSYQGRVYPRPGMCPSTGFTFSSTKDFPDVYIOFARNHPL 407

OY 311 DK--V--F--RTSSL-K--GYHSSLNPRPGKCLPQO-QPIP-TETP--QVAD--R-HPE 354

Db 408 MYNSVLPTGG-R-PLFLOYGANTYTTQIADRYAADAQH-YDVLFTGIDVGTVLKIVISP 464

OY 355 VAQREPMRPLKPLP-H--SKYHQKVAVHRMQASHGETFHVLLTTRGTIRHKVVE-P 410

Db 465 KRSRSABGLLEELHVEDSAVTSMTQSKRHOLYVSRSAVQIALHRCAAHGRVCT 524

OY 411 -GEQHSFAFNIMEIOPRRRAAIQTMSLDAERKRLYSSQMEVSQVPLDCEYVGGCH 469

Db 525 ECCLARDPYCAMDGYACTRFOPSAKRFRRODVNRGDPSTLCSGSSRPALLEHKKVGE 584

OY 470 GCLMSRDPYCGMDQRCISITSSERSVL--QGINPAEPHKECPNPKRPAKL-QKVS LAP 526

Db 565 GSSAFLECEPRSLQARVETFORAGV 610

OY 527 NSRYLSCPMESRHATYSWR-HKENV 551

RESULT 7

ID Q93018 PRELIMINARY: PRT: 750 AA.

AC Q93018;

DT 01-FEB-1997 (TREMBLREL. 02, CREATED)

DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE SEMAPHORIN V.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RA DANTE M., WAMSLLEY P.;

RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RA WATERSTON R.;

RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]

RP SEQUENCE FROM N.A.

RA WATERSTON R.;

RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; U73167; G2880035; -.

PFAM; PF00047; 19; 1.

SO SEQUENCE 750 AA; 83034 MW; C8C87E34 CRC32;

Query Match 16.5%; Score 716; DB 4; Length 750;
Best Local Similarity 33.0%; Pred. No. 1,886-148;
Matches 167; Conservative 115; Mismatches 170; Indels 54; Gaps 41;

Db 113 EGMNFVKLLHAYNRTHLACGTGAHPHCAPVEVGHRAEPEVRLRDPGRIDGKSPYD 172

OY 93 DCENITTLER-RSEGLLACGTINARHPS-C-WNLV-N-G--TVVPL--G--EM-RGIAPFS 141

Db 173 PRHRAASVLYGELYSGVA-ADLMGRDFTIFRSLGQRPSLRTPEHDSRWLNPEKVFYEW 231

OY 142 PDENSLVLEGEDEVYSTIRKQEVNKG-IPRFRIRGESELYT-S-DTV-MQNPQFIKAT- 196

Db 231 IPESNPDDDKTYFFRFRTEAVALRSLYSRNGQICRNDVGGORSL-VNKTTFELKA 290

OY 197 IVHODQAYDDKTYFFREDNPDKNPE-APLNVSRVAQICRDOGESSLSYSKWNTEFLKA 255

Db 291 RLVCSPVGEQDTHFDOLDFLLS-SRDH-RTPLLYAVFSTSSIFQCSAVCYVSMNDV 348

OY 256 MLVCS--DAATKNKRNRLQDVFLLRDPSCGMROTIRYGVFSNPMN-Y--SAVCYISLGD 310

Db 349 RRAFLGPRANKEGPMHGVSYQGRVYPRPGMCPSTGFTFSSTKDFPDVYIOFARNHPL 408

OY 311 DK--V--F--RTSSL-K--GYHSSLNPRPGKCLPQO-QPIP-TETP--QVAD--R-HPE 354

Db 409 MYNSVLPTGG-R-PLFLOYGANTYTTQIADRYAADAQH-YDVLFTGIDVGTVLKIVISP 465

OY 355 VAQREPMRPLKPLP-H--SKYHQKVAVHRMQASHGETFHVLLTTRGTIRHKVVE-P 410

Db 466 KRSRSABGLLEELHVEDSAVTSMTQSKRHOLYVSRSAVQIALHRCAAHGRVCT 525

OY 411 -GEQHSFAFNIMEIOPRRRAAIQTMSLDAERKRLYSSQMEVSQVPLDCEYVGGCH 469

Db 526 ECCLARDPYCAMDGYACTRFOPSAKRFRRODVNRGDPSTLCSGSSRPALLEHKKVGE 585

OY 470 GCLMSRDPYCGMDQRCISITSSERSVL--QGINPAEPHKECPNPKRPAKL-QKVS LAP 526

Db 566 GSSAFLECEPRSLQARVETFORAGV 611

OY 527 NSRYLSCPMESRHATYSWR-HKENV 551

RESULT 8

ID Q99985 PRELIMINARY: PRT: 751 AA.

AC Q99985;

Db 346 DIRVNGPFAHKKEGPNYQMPFGSKMPYPBPGTCTPGTTPSKMSTVDYDDEVINEMRS 405
QY 309 DIDKVFRTS-SLK-G-YH---SS-LPNPRGKCLPDQ-QP-I-PTETF--QVAD--R- 351
Db 406 HPLMQAVYPLQRRPLVVRT-GAPYRLTTIANDOVDSADG-RYEVLFCTDGGTQOKYIV 463
QY 352 HPEVAQREVPBG--PLKTPLEHRSKYHOKAVAHRMQASHGETHFHLYLTDDRTIHKV 409
Db 464 LPKDOEVEELMEVEVEFKDPAPKTMITISSKROOLYVASAVGTHLSLRQAYGAAC 523
QY 410 -PGEDEHFAFNINKEIOFFRRRAAIIQTMSLDAERKRLYSSQMEVSQVPLDCEYGGGC 468
Db 524 ADCCLARDPYCAMDQACSRYTASSRRSRQDVHGNPIROCCFGNSNANKNAVESYQ 583
QY 469 HGLCMSRDPYCGMDQGRCS-ITYSSER-SVLQSIINPAEPHKCP--NPKPKAPLQKVS 524
Db 584 GVAGSAATLECOAPRSPOATVKW 605
QY 525 A-PNSRYLLSCPMESRHATYSW 545

RESULT 11
ID 088633 PRELIMINARY: PRT: 754 AA.
AC 088633:
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV ISOFORM A.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SCUROGNATHI; MORIDAE; MURINAE; MUS.
RN [1]
RA SEQUENCE FROM N.A.
RA ECKHARDT F., MEYERHANS A.;
RT "Molecular cloning and expression pattern of a murine semaphorin
RT homologous to H-scra iv."
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF080091; G3377768; -.
SQ SEQUENCE 754 AA; 85133 MW; EC05BDF3 CRC32;

Query Match 15.7%; Score 682; DB 11; Length 754;
Best Local Similarity 31.8%; Pred. No. 1,52e-119;
Matches 141; Conservative 119; Mismatches 140; Indels 43; Gaps 31;

Db 171 PYDKLDTASALINELIAGYVI-DEMGTDAAIFRTLGKOTAMRTDQYNSRWLNDSFEIH 229
QY 139 PESPDENSLVFEQDEYVSTIRKQYNGK-IPRFRRIRGESELYTS--DTV-MQNPQFIK 194
Db 220 AELLPDSAEKNDKLYFFFRERSAE-APQNPAYVARIGRICLNDGHCCL-VNKMSTFL 287
QY 195 ATIVHOD-QAVDDKLYFFREDNPDKPEAPLNVSRVAQLCRGDGGSLSVSKMNTFL 253
Db 288 KARLVCSVPGEGCTHDELDQVFN-QQTOOI-RNPVIYANFTSSGVSFRSAYCAVYSM 345
QY 254 KAMLYCS--DAATNKNNRLQDVFLLPDSGQMPDTRYIGVFSNP-WNY--SAVCVSTL 307
Db 346 ADIRAVENGPEFAHKEGPNYQMPFGSKMPYPBPGTCTPGTTPSKMSTKDYDDEVINEMR 405
QY 308 GDIDKVFRTS-SLK-G-YH---SS-LPNPRGKCLPDQ-QP-I-PTETF--QVAD--R 351
Db 406 THPLMQAVYPLQRRPLVVRT-GAPYRLTTIANDOVDAADG-RYEVLFCTDGGTQOKYIV 463
QY 352 -HPEVAQREVPBG--PLKTPLEHRSKYHOKAVAHRMQASHGETHFHLYLTDDRTIHKV 408
Db 464 VLPRKDOEVEELMEVEVEFKDPAPKTMITISSKROOLYVASAVGTHLSLRQAYGAAC 523
QY 409 E-PGEDEHFAFNINKEIOFFRRRAAIIQTMSLDAERKRLYSSQMEVSQVPLDCEYGGG 467
Db 524 CADCLARDPYCAMDQACSRYTASSRRSRQDVHGNPIROCCFGNSNANKNAVESQ 583
QY 468 HGLCMSRDPYCGMDQGRCS-ITYSSER-SVLQSIINPAEPHKCP--NPKPKAPLQKVS 523
Db 584 YGVAGSAATLECOAPRSPOATVKW 606

QY 524 LA-PNSRYLLSCPMESRHATYSW 545

RESULT 12
ID 090607 PRELIMINARY: PRT: 772 AA.
AC 090607:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COLLASIN.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RA SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE; 94006554.
RA LDO Y., RAIBLE D., RAPER J.A.;
RT "Collapsin: a protein in brain that induces the collapse and
RT paralysis of neuronal growth cones."
RL CELL 75:217-227(1993).
DR EMBL: U02528; G410079; -.
DR PIRAM; PF00047; Iq; 1.
SQ SEQUENCE 772 AA; 88867 MW; BC6CB9E8 CRC32;

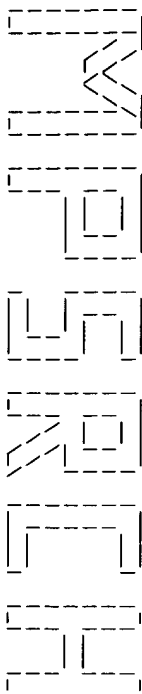
Query Match 15.7%; Score 681; DB 13; Length 772;
Best Local Similarity 31.5%; Pred. No. 2,77e-139;
Matches 146; Conservative 119; Mismatches 150; Indels 48; Gaps 38;

Db 166 RKSHPYDKLLTASLVDELGYSTA-ADFKGRDPAIFRLGHHNPIRTEOHDSRWLNDP 224
QY 135 RGVAPFSPDENSLVLFEGDEYVSTIRKQYNGK-IPRFRRIRGESELYTS--DTV-MQNP 190
Db 225 RFIASHLIPESDNEDDKYFFERENAIIDGHTCKATHARIGQICKNDPGHREL-VNKM 283
QY 191 QFIKATYVHO-DOAYDDKLYFFREDNPDKPEAPLNVSRVAQLCRGDGGSLSVSKM 249
Db 284 TTFPLKARLCSVPGPNIGDTHFDELQVFLM-NSKDP-KNPYVGVTTSSNIFKGSAYC 341
QY 250 NTFPLKAMLYCS--DA-ATNKFNRLQDVFLLPDSGQMPDTRYIGVFSNPW-N--SAVC 303
Db 342 MYSMTDVARNVFLGYAHROGPNYQMPVYQGVYPRPGTCSKTFGGFSDTKLDPEVIT 401
QY 304 VYSLGDIDKVFRTS-SLK-G-YH--S-S-LPNPRGKCLPDQ-QP-I-PTETF--QV-A 349
Db 402 FARSHPAVNNVFPINS-RPIMITDVQYQTOVLVDRVADBDQO-YDVWFICGTDIGTLV 459
QY 350 -DR-HPEVAQREVPBGPLKTPLEHRSK--YHOKAVAHRMQASHGETHFHLYLTDDRTIH 405
Db 460 KVSIPKETWLEBEVLEEMTVFREPTVISAMKISTKQOOLYGSATGVSQLPLHRCDV 519
QY 406 KVVE-PGEDEHFAFNINKE-IQPFRRRAAIIQTMSLDAERKRLYSSQMEVSQVPLDCEV 463
Db 520 YGKACAECCLARDPYCAMDSSGSRYPYAKRTRTODIRNGDPLTCSDLQHHNDPSQ 579
QY 464 YGCGHGLCMSRDPYCGMDQGRCSISYSS-RSVL-QSINPAEPHKCPN-PKPKAP-- 518
Db 580 TLEKTIIGVENSSTFLECSFKSORATYV-WQFQKQNDHKV 621
QY 519 -LQ-KVSLA-PNSRYLLSC-PMESRHATYSWR-KRENEVQSC 556

RESULT 13
ID 088632 PRELIMINARY: PRT: 785 AA.
AC 088632:
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV ISOFORM B.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SCUROGNATHI; MORIDAE; MURINAE; MUS.


```
QY      195  ATIYHOD-QAYDDKLYFFREDNDPKNEAPLNVSRVAQLCRGDGGESSLSVSKMNTFL 253
      | : : : : | | : | | | | : : : | : : : : | : | | | : | | : | | |
Db      319  KARLVCSYVGEDEGIEHDELODFV-OOTODV-RNPVIYAVETSSGSVFRGSAVCYXSM 376
      | | | | | : : : : : | | | | : : : | | : : : : | | | | | :
QY      254  KAMLYCS---DAATKKNFNRLQDFVLLPDPGQWMDTRVYGVFSNP-WNY--SAYCVYSL 307
      : | : | : : | : | : : | : | : | : | : | : | : | : | : | :
Db      377  ADIRNVFNGPFAHKEGPNYQWMPESGKMPYPRPGTCPGGTFTPSMKSTKDYPPDEVINFMR 436
      : | : | : : | : | : : | : | : | : | : | : | : | : | : | :
QY      308  GDIDKVFRTS-SLK-G---YH-----SS-LPNPRPGKCLPDQ-QP-I-PTETF--QVAD--R 351
      : | : | : : | : | : : | : | : | : | : | : | : | : | : | :
Db      437  SHPLMYQAVYPLQRRRLVVRT-GARYRLTTIYAVQVDAGDG-RYEVFLGTDRTGVQKI 494
      | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      352  -HPEVAQRVEPVG--PLKTPLEHSHKYHQKVAVHRMQASHGETFHVLYLTTDRGTIHKVY 408
      | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      495  VLPKDDQEMEELMEVEVFKDPAPVKTMTITSSKROOLYVASAVGVTHLSLHRCQAYGAA 554
      | : : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
QY      409  E-PGEQEHSAFNIMEIOPFRRAAIOTMSLDAERKLIYSSQWESQVPLDLCEVYGGG 467
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      555  CADCCCLADPYCAMDGQACSRYTASSKRRSRQDVRHGNPIRQCRGFNSNANKNAVESVQ 614
      | : | : | | | : | | : | : | : | : | : | : | : | : | : | :
QY      468  CHGCLMSRDPYCGWMDQGRGIS-IYSSESR-SYLQSIINPAEPHKECP--NPKPDKAPLQKYS 523
      | : | : | | | : | | : | : | : | : | : | : | : | : | : | :
Db      615  YGVAGSAFAFLECGPQRSQATYKW 637
      : : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      524  LA-PNSRYLLSCPMESRHATYSW 545
      : : | : | : | : | : | : | : | : | : | : | : | : | : | :
```

Search completed: Thu Jul 8 19:19:45 1999
Job time : 54 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:25:02 1999; Maspar time 23.11 Seconds

Tabular output not generated. 522.665 Million cell updates/sec

Title: >US-09-041-236-2
(39-606) from US09041236.Pep (34 of 45)

Perfect Score: 4236
Sequence: 1 QTEPHYLVFHEPSSSWVG.....SYFNAQHWOLLPEGIMAE 568

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.340; Variance 153.758; scale 0.236

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	661	15.6	771	13	R71380	Human semaphorin III	6.72e-51
2	620	14.6	477	13	R74175	Human collapsin.	6.70e-47
3	608	14.4	775	33	W63748	Human semaphorin.	9.86e-46
4	606	14.3	861	22	W17658	Mouse CD100 antigen.	1.54e-45
5	599	14.1	861	32	W58540	Human semaphorin.	7.39e-45
6	544	12.8	862	22	W17657	Human CD100 antigen.	1.58e-39
7	510	12.0	776	32	W51313	Rat semaphorin W.	3.01e-36
8	463	10.9	441	13	R71381	Vaccinia virus semaphorin	5.07e-32
9	445	10.5	730	13	R71379	Grasshopper semaphorin	2.10e-28
10	428	10.1	587	32	W51314	Human semaphorin W.	8.31e-25
11	390	9.2	974	33	W64221	Human secreted protein	1.12e-23
12	378	8.9	712	13	R71384	Tribolium semaphorin	1.21e-22
13	367	8.7	650	13	R71382	Drosophila semaphorin	3.56e-22
14	362	8.5	724	13	R71383	Drosophila semaphorin Y.	3.07e-21
15	352	8.3	930	32	W57260	Human semaphorin Z.	9.96e-19
16	325	7.7	888	25	W19857	Human semaphorin Z.	9.96e-19

17	316	7.5	887	25	W19856	Rat semaphorin Z.	6.78e-18
18	296	7.0	929	32	W57259	Rat semaphorin Y.	4.69e-16
19	157	3.7	122	13	R71385	Varicella major virus S	7.87e-04
20	101	2.4	2329	25	W25038	Partial BRCA2 cancer f	1.78e-01
21	99	2.3	591	4	R23006	Protein transcribed f	2.47e+01
22	98	2.3	832	33	W61092	Tag DNA polymerase I	2.91e+01
23	96	2.3	832	33	W61090	Tag DNA polymerase I	4.02e+01
24	92	2.2	832	33	W61091	Tag DNA polymerase I	7.64e+01
25	92	2.2	854	30	W56309	Claas II S-receptor K1	7.64e+01
26	92	2.2	855	6	R29815	S receptor kinase pro	7.64e+01
27	92	2.2	4572	30	W52845	A. mediterranei r1fam	8.95e+01
28	91	2.1	478	1	R04881	Recombinant elastase.	8.95e+01
29	89	2.1	534	25	W25031	Partial BRCA2 cancer	1.23e+02
30	89	2.1	554	13	R65209	Novel thermostable DN	1.23e+02
31	91	2.1	598	9	R48631	Sequence of nuclear r	8.95e+01
32	90	2.1	638	5	R03924	E. coli HSP (dnak).	1.05e+02
33	89	2.1	680	4	R23143	Mutant thermostable D	1.23e+02
34	89	2.1	832	13	R76593	DNA-polymerase (F73L,	1.23e+02
35	89	2.1	832	1	P90556	Purified native therm	1.23e+02
36	89	2.1	832	33	W61087	Tag DNA polymerase I	1.23e+02
37	89	2.1	832	33	W61088	Tag DNA polymerase I	1.23e+02
38	89	2.1	832	13	R76590	Tag DNA-polymerase RE	1.23e+02
39	89	2.1	832	33	W61089	Tag DNA polymerase I	1.23e+02
40	89	2.1	833	36	W59942	Amino acid sequence o	1.23e+02
41	89	2.1	833	27	W24211	Cleavage DN nuclease.	1.23e+02
42	89	2.1	833	27	W24212	Cleavage DA nuclease.	1.23e+02
43	89	2.1	1242	30	W52287	Rattus norvegicus cdo	1.05e+02
44	89	2.1	3418	26	W19211	Human breast cancer s	1.23e+02
45	89	2.1	3418	26	W23287	Human breast and ovar	1.23e+02

ALIGNMENTS

RESULT 1
ID R71380 standard; Protein: 771 AA.
AC R71380;
DR 21-NOV-1995 (first entry)
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW varicella major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN M09507706-A.
PD 23-MAR-1995.
PE 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB: Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2: Page 60-63: 101pp: English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (087441), human semaphorin III,
CC vaccinia virus semaphorin IV (087443), Drosophila semaphorin I and II
CC (087444-5), Tribolium semaphorin I (087446) or varicella major (smallpox)
CC virus semaphorin IV (087447) genes were used to generate a series of
CC peptides (870370-870418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA:
Query Match 15.6%; Score 661; DB 13; Length 771;
Best Local Similarity 30.3%; Pred. No. 6.72e-51;
Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;
Db 166 rtkspdyphklltssllldelysgta-adfmgdrfdiftlghhphirceghdsrwlnp 224

Qy	135	RGYAPFSDEMSLVLFEEDEYVYITRQOEYNGK-IRPFRRIRGSESLYTS-DTV-MQNP	190
Db	225	kfiashlisesdapedkvylffifenaIdgehsqalharigqlckndfigghysl-vnwk	283
Oy	191	OFIKATIVHO-DOAYDRIKYFFREDDNPKNEAPLWVSRVAOLCRDQOGESSLSYSKW	249
Db	284	ctfllkarlicsvpprngldthfcdlgvflm-nkdp-knpvygvttesnlfkgsawc	341
Oy	250	NTEFLKAMLVCS-DA-ATNKNFNLADVFLLRPDSQGMROTJRYGVGFSNMWN-Y--SANC	303
Db	342	mysmedvrrvrlgpyahndgpnvgwvyqgrvprprrptocpsktffgfsdtkdldpdvlt	401
Oy	304	VYSLGDIDIKVERTS-SLK-G--YH--S--S-LPMPRRGKLLPDO-QPI-PTFFF-QV-A	349
Db	402	farsphamyprvfmmnrprilyktdvnyftqglvdi.vdaedsg-ydvmfigtdvltvik	460
Oy	350	-DR-HPEVAQVBERMGPRLKTPRL-FHSKYHXOKAVHBMQASHGETHYLVLTDRGRIHK	406
Db	461	vyslprkewyldeevlleemlvfrefrcalisesmestqgqlylgystgqvqdlrlnhcdly	520
Oy	407	VVE-PGEOEHSAFENIME-IDPFRRAALITMSIDARRKLYVSSOWESQVPLDCEVY	464
Db	521	gkacaeccclardpycawdasgacsyufctakrrtrrgdrlngdoltshesdihhnhghnsp	580
Oy	465	GGCGHGLMSMDPRYCGMDQGCISITYSE-RSVL-QSINAPERHKECPNKKPRK----	AP 518
Db	561	eerlllygvensctflescpksqalvywqfqrtnearkeelrvvdhllirtdgqlllrsjl	640
Oy	519	LQKYSLA-PNSRYVLSC-PMESRATY-SW-RHKEVY-EQSCERPCHO-SPCILFIENLT	572
Db	641	qkdsngylchavvehfig 658	
Oy	573	AQOYGHVFCENQEGSYFR 590	

RESULT	2	
ID	R7A175	standard; Protein: 477 AA.
AC	R7A175;	
DT	01-NOV-1995	(first entry)
DE	Human collapsin.	
KW	Collapsin; antibody; therapy.	
OS	Homo sapiens.	
FT	Key	Location/Qualifiers
FT	binding_site	9..19
FT		/note, "antibody binding site"
FT	binding_site	51..65
FT		/note, "antibody binding site"
PN	US5416197-A.	
PD	16-MAY-1995.	
PE	15-OCT-1993.	136922.
PR	15-OCT-1993:	US-136922.
PA	(TYPE-) UNIV PENNSYLVANIA.	
P1	Luo Y, Raper JA:	
DR	WPI: 95-193478/25.	
DR	NPDB: 092331.	
PT	New antibody to human collapsin - used to inhibit the activity of	
PT	collapsin, to induce neurite out-growth and to treat individuals with	
PT	nerve damage.	
PS	Claim 2; Columns 15-18; 11pp; English.	
CC	An antibody capable of specifically binding at least a portion of	
CC	the collapsin protein can be used to purify human collapsin and	
CC	to inhibit the activity of the protein. It can be used to induce	
CC	neurite outgrowth by neuronal cells and to treat individuals	
CC	suffering from nerve damage.	
Sequence	477 AA:	

Query Match 14.6%; Score 620; DB 13; Length 477;
Best Local Similarity 31.2%; Pred. No. 6,70e-47;
Matches 134; Conservative 110; Mismatches 145; Indels 41; Gaps 32;

Oy	198	VHODQAYDCKIYEFREDDNPDKNEAPRLNWSRVQOLCRDGGGSSLSVSKMNTFLKAML	257
Db	60	lcsvpprgnltthdeldvflm-nfkdp-knpvuyvfttsnltfkysavcmysmdvr	1177
Oy	258	VCS--DA-ATKKNFNRLDDELDPDPSGOWMDTVYGVFSNPNW-Y--SAWCVYSLDID	3111
Db	118	rvf1gryahidgpmvuywrygyrpyrprvcspskctfsgtastckldbdvfttarshpm	1777
Oy	312	KVFRIS-SLK-G--YH--S--S-LPNRRPKCLPDQ-QPT-PLTF--QV-A-DR-NREV	3555
Db	178	yprvfpmnrplvltktdnyvftglvdydaedqg-ydmf1gtdvtylkwvs1pck	2366
Oy	356	AQRVPMRPLKTP-L-FHSKYHYQKVAVHBMQASNGELFNHLYLTDRGITHKVE--PGEQ	4133
Db	237	wydlleevl1eemtyfrepbtasame1stkgqly1gstagaq1p1hrcdiygtakaecc	2366
Oy	414	EHSFNF1ME-IOEFPRRMAA1QOTSLDAERKLTVSSOWEVSQVPLDICEVYGGGCHGL	4722
Db	297	lar0pawd6sacs1yfrtakrttrtdg1ngbdlhcsdlhndhbnhspeor1lyv	3566
Oy	473	MSRDPYCGMDGRCIS1YSSE-RSVL-QSINPABPHKECPKPKDK---APLQKVS-LA	5233
Db	357	enstf1leacspdksgataylvwqfgrneerkeelrvdahl1rtddg1llrs1qkdsngyl	4166
Oy	526	PNSRYULSC-PMESHARY-SW-NHKENV-EQSCPEHQ-SPNCLITENLTAQYCHYF	5800
Db	417	chavengftq 426	
Oy	581	CEAQEGSYFR 590	

RESULT 3
ID W63748 standard: Protein: 775 AA.
AC W63748;
DT 01-OCT-1998 (first entry)
DE Human semaphorin.
KW Human; semaphorin; diagnosis; nerve disease; nerve growth inhibitor;
KW neurological disease; atopic skin inflammation; autoimmune disease;
KW pain.
OS Homo sapiens.
PN W09822504-AL.
PD 28-MAY-1998.
PE 12-NOV-1997; J04111.
PF 15-NOV-1996; JP-321068.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Furuyama T, Inagaki S;
PI WPI: 98-312416/27.
DR N-PSDB: V35367.
PT Gene encoding new semaphorin nerve growth inhibitor - useful in
PT diagnosis, treatment and study of neurological diseases
PS Claim 1: Page 33-37; 49pp: Japanese.
CC The present sequence represents human semaphorin, a nerve growth
CC inhibitor. The semaphorin protein, and gene encoding the protein,
CC and their derivatives, are used in the diagnosis, treatment and
CC study of neurological disorders such as atopic skin inflammation,
CC autoimmune diseases and pain.
SQ Sequence 775 AA;

	Query Match	14.4%	Score 608:	DB 33:	Length 775:
	Best Local Similarity	32.1%	Pred. No. 9,86e-46:		
	Matches 125:	Conservative	98:	Mismatches 130:	Indels 36: Gaps 28:
Db	167	rgrcpfdbpnsfstvslvgnelfaglyls-dwgrtsafmsgrkighrttendderllkcp	225		
Oy	135	RGYAPFSDEMSLVLFESDEEYSTRROEYNGK-IPFRFR-I-R-GE-SELYSTSTVWQNP	190		
Db	226	kfvgsymipdnedcdnkmyfftekaieaenahitlytrgvlcvndmgqrll-vnk	284		
Oy	191	QEKATIVHQDQAVDD-KIYYFREDPNDDKNPEAPLVNVSRAQCGRDGGESLSVSKW	249		
Db	285	stflkarlivespmmgldtyfdeledvflrp-trdp-knpulfglntnsnlfgrhvac	342		
Oy	250	NTFLKAMLVSCDAATN-KN-FNLSLOVFLFLPDSGCMROTRVYGVSPN-V-S-AVC	303		

ID	Accession	Description	Location/Qualifiers
Db	343	Wymssiireafngnyahkegepeylwsllyegkvypyrpagscaskvnggkkygttkypddai	
Qy	304	VYSGLDIDKVERTS-SLK-G--YHSSL-----PNPRREKCLPD-QQPI--PTETP-OVAD	
Db	403	rfaamhrlmyapllrvphvkbpillvctdgyunrlqavldvaeedqg-ydvlfifgdtglvl	
Qy	351	RHPEVAQGV-EPKGPL-KTP-LFHS--KYHGXKAVHVMQASHETHVHLTLTTDRITIH	
Db	462	kvitlyngeteemeevllleelqfkdapalismsaiskrqgllygsasavaqvrffhcdm	
Qy	406	KVPEEGEHEHFAFN-IME-IQPRRAAAIOTMSLDIAERRLKLVYSOWEYVSQVPLDCEV	
Db	522	ygsacadcclardpycawdgiscrrypt 550	
Qy	464	YGGGCHGCLMSRDPYCGMDGRCTIYSS 492	
RESULT	4		
ID	W17658	standard: protein: 861 AA.	
AC	W17658	24-JUL-1997 (first entry)	
DE	Mouse CD100 antigen.		
KM	CD100 antigen: semaphorin: leukocyte: B cell; T cell; lymphocyte;		
KM	vaccine.		
OS	Mus sp.		
FH	key	Location/Qualifiers	
FT	peptide	1..41	
FT	protein	/label= sig_peptide	
FT	domain	42..861	
FT	domain	/label= Mat_protein	
FT	domain	42..553	
FT	domain	/label= Semaphorin_domain	
FT	domain	534..630	
FT	domain	/label= Ig-like_domain	
FT	domain	631..732	
FT	domain	/label= Stalk_domain	
FT	domain	734..752	
FT	domain	/label= Transmembrane_domain	
FT	domain	753..861	
FT	domain	/label= Cytoplasmic_domain	
FT	modified_site	807..814	
FT	modified_site	/label= Phosphorylation	
FT	modified_site	/note= "putative tyrosine phosphorylation site"	
PN	W09717368-A1.		
PD	15-MAY-1997.		
PF	12-NOV-1996: U18645.		
PR	09-NOV-1995: US-556422.		
PA	(DAND) DANA FARBBER CANCER INST.		
PI	Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;		
DR	WPI: 97-280982/25.		
DR	N-PSDS: T60666.		
PT	Nucleic acid molecule encoding CD100 antigen - which stimulates		
PT	leukocyte response, e.g. B cell aggregation, differentiation,		
PT	survival and T cell proliferation		
PS	Example 8: Page 86-89: 135pp: English.		
CC	Mouse CD100 antigen (W17657) is a novel leukocyte semaphorin-like		
CC	protein that stimulates a leukocyte response, including B cell		
CC	aggregation, B cell differentiation, B cell survival and/or T cell		
CC	proliferation. Its amino acid sequence was deduced from a cDNA		
CC	clone (T60666) isolated from murine T cells. Human CD100 antigen		
CC	(W17657) has also been identified. CD100 polypeptides and fusion		
CC	proteins, nucleic acids, and host cells expressing CD100 can be		
CC	utilised in diagnostic and therapeutic methods involving modulation		
CC	of B and T cell responses, neuron axonal growth and immune cell-		
CC	neve cell interaction.		
CC	Sequence 861 AA:		
Qy	Query Match	14.3%: Score 606; DB 22; Length 861;	
Qy	Best Local Similarity	32.4%: Pred. NO. 1,54e-45;	
Qy	Matches 143; Conservativeness 106; Mismatches 145; Indels 47; Gaps 31;		
Qy	107 eclnyirvrlgplststlyvcgtnafgptcdhlntltfklfgsedgkrpcfdpahyts 166		

QY	93	DCENTITLLER-RSEGLIACCTNRRHSC--WNLVNGTVVPLGCM-RGIAPFSDENSLV	148
Db	167	vmvsggelysgts-ynfligsepilsrnshsplrteyailpwnpsfxfadviqspdre	225
QY	149	LFEQGEVYSTIRKQEVNKNKIPRRIRIGESBLYTSDFV--MNPQFIKATIVHQ--D-Q	202
Db	226	geddkvvyffifvevsvefvlklnprvarvcksgdgglttlq-kkxtsfikarlicskp	284
QY	203	AYDDKIYFFEFEDNDPKRPEAPLWVSRAQICRSDGEGESLSVSKWNTFLKMLWCSDA	262
Db	285	dsqlyfnllgdvfyliar-qf-kbpvfyavftplnnvgjlsavcaelylaeevafsrqky	342
QY	263	ATNKNFENLQDVEFLPPDSGOMRQTRRYGCFSPNWNV---SAYCVISLGDIDKYF-R---	315
Db	343	mqsatvegshkwtvryngpvpctprgacisdearaanyslmpcktlqfvkxhplmd	402
QY	316	--TSSL-----K-G-YHSSLPNRPBKCL-PDQQPIP-TENFGVADRNPE-V-AQR-VE-	360
Db	403	svtprldmrplkikdvnvgtvlvotrtgalqgtfdvfmfistdgalbkavylkcvh-v	460
QY	361	PMGPRKT-P-LFHSKYNHQKVAVRMOASHGETFHVLYLTTRDTIRHKVVEPGEQESFA	418
Db	461	--leetgfifdsepylltllsskgrkfyvagsnsgvqparlafcehgs--cedaviard	517
QY	419	FNIMIDQFRRAAAIQIMSLDAER-RKL-YVSSQWEVSQVPLDLCEYVGGGCHCCLMSRD	476
Db	518	pycawspalkacvtllhgeas	538
QY	477	RYCGWMDQ-R-CISIVSERS	495
RESULT	5		
ID	W58540	standard; Protein; 861 AA.	
AC	W58540.		
DT	02-SEP-1998	(first entry)	
DE	Human semaphorin.		
KM	Human; semaphorin; diagnosis; nervous disease; immune disease;		
KN	nerve extension inhibiting activity.		
OS	Homo sapiens.		
PN	J10155490-A.		
PD	16-JUN-1998.		
PF	27-NOV-1996; 332900.		
PR	27-NOV-1986; JP-332900.		
PA	(SUMU) SUMITOMO SEIYAKU K.		
DR	WPI: 98-391044/34.		
DR	N-PSDB: V31121.		
PT	New human semapholin gene - useful in the diagnosis of nervous system		
PI	and immune disorders		
PS	Claim 1: Page 10-12; 15pp; Japanese.		
CC	The present sequence represents human semaphorin (translated from the		
CC	Japanese specification as semapholin). Semaphorin has nerve extension		
CC	inhibiting activity. The semaphorin gene is useful for the diagnosis,		
CC	treatment and researches on nervous diseases and immune diseases.		
SEQ	Sequence	861 AA;	
Query Match	14.1%.	Score 599; DB 32; Length 861;	
Best Local Similarity	32.4%.	Pred. No. 7.39e-45;	
Matches	143;	Conservative 105; Mismatches 146; Indels 47; Gaps 31;	
Db	107	eclynfivrlglsstslvscgtngafpctdhlnltfskffgsgedgkyrcpfdpahys	166
QY	93	DCENTITLLER-RSEGLIACCTNRRHSC--WNLVNGTVVPLGCM-RGIAPFSDENSLV	148
Db	167	vmvsggelysgts-ynfligsepilsrnshsplrteyailpwnpsfxfadviqspdre	225
QY	149	LFEQGEVYSTIRKQEVNKNKIPRRIRIGESBLYTSDFV--MNPQFIKATIVHQ--D-Q	202
Db	226	geddkvvyffifvevsvefvlklnprvarvcksgdgglttlq-kkxtsfikarlicskp	284
QY	203	AYDDKIYFFEFEDNDPKRPEAPLWVSRAQICRSDGEGESLSVSKWNTFLKMLWCSDA	262
Db	285	dsqlyfnllgdvfyliar-qf-kbpvfyavftplnnvgjlsavcaelylaeevafsrqky	342
QY	263	ATNKNFENLQDVEFLPPDSGOMRQTRRYGCFSPNWNV---SAYCVISLGDIDKYF-R---	315
Db	343	mqsatvegshkwtvryngpvpctprgacisdearaanyslmpcktlqfvkxhplmd	402
QY	316	--TSSL-----K-G-YHSSLPNRPBKCL-PDQQPIP-TENFGVADRNPE-V-AQR-VE-	360
Db	403	svtprldmrplkikdvnvgtvlvotrtgalqgtfdvfmfistdgalbkavylkcvh-v	460
QY	361	PMGPRKT-P-LFHSKYNHQKVAVRMOASHGETFHVLYLTTRDTIRHKVVEPGEQESFA	418
Db	461	--leetgfifdsepylltllsskgrkfyvagsnsgvqparlafcehgs--cedaviard	517
QY	419	FNIMIDQFRRAAAIQIMSLDAER-RKL-YVSSQWEVSQVPLDLCEYVGGGCHCCLMSRD	476
Db	518	pycawspalkacvtllhgeas	538
QY	477	RYCGWMDQ-R-CISIVSERS	495

[illegible]

RESULT	6
ID	Wt17657 standard; Protein; 862 AA.
AC	Wt17657;
DT	24-JUL-1997 (first entry)
DE	Human CD100 antigen.
KW	CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte; vaccine.
OS	Homo sapiens.
FH	Key
FT	peptide
FT	/label/ Sig-peptide
FT	protein
FT	/label/ Mat-protein
FT	/label/ Ig-like_domain
FT	/label/ Semaphorin_domain
FT	/label/ Ig-like_domain
FT	/label/ Stalk_domain
FT	/label/ Transmembrane_domain
FT	/label/ Cytoplasmic_domain
FT	/label/ Phosphorylation
FT	/note/ "putative tyrosine phosphorylation site"
FN	modified_site
PD	15-MAY-1997
PF	12-NOV-1996; U18645
PR	09-NOV-1995; US-556422.
PA	(DAND) DANA FARBER CANCER INST.
PI	Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL, WPl: 97-280982/25.
DR	N-PSSD; T60665.
PT	Nucleic acid molecule encoding CD100 antigen - which stimulates leukocyte response, e.g. B cell aggregation, differentiation, survival and T cell proliferation
PS	Claim 7: Page 70-72; 135pb: English.
CC	Human CD100 antigen (Wt17657) is a novel leukocyte semaphorin-like protein that stimulates a leukocyte response, including B cell aggregation, B cell differentiation, B cell survival and/or T cell proliferation. Its amino acid sequence was deduced from a cDNA clone (T60665) isolated by COS cell expression cloning. CD100 polypeptides and fusion proteins, nucleic acids, and host cells expressing CD100 can be utilized in diagnostic and therapeutic methods involving modulation of B and T cell responses, neuron axonal growth and immune cell-nerve cell interaction.
SQ	Sequence 862 AA:
Query Match	12.8%; Score 544; DB 22; Length 862;
Best Local Similarity	29.3%; Pred. NO. 1,58e-35;
Matches 157; Conservative 134; Mismatches 189; Indels 55; Gaps 39;	

```

Db 107 echnlvfvlprlpaaetslycgcnaafgpadchllsfcflknedgagrcpfdpabays 166
Oy 93 DCENYITLLRRRE-GULACGTNARHPSCMWLVNCTVYPLG--EM-RGIAPFSDEN-SL 147
Db 167 vvmvdg-elyegts-yufligseplrsnsehplrtfeyalpwlnepstvfadavlrkspdp 224
Oy 148 VLFEGGDEVYSTIKROENKGRIPFRRIRESSELYTSDTV--NONPOFIKATIVAO--D- 201
Db 225 dgeddgyrfftevsvegefvlvrlprlarckbdgggltdlg-kwstflkrlcstr 283
Oy 202 QAVDDKTYITFFREDNDPKKNDEARLVNSRPAQLCRDQGGESSLSYSKNTFLKMLVCSO 261
Db 284 pdsqglfvlrvdlfvvlrtp-gl-kvrvfyalfplglnvglsavcaaynlstaeevfhqk 341
Oy 262 AATKNKNRRLQDVFLLRDPDSGQWRDTRGVGFSNPWNT---SAYCVYSLGIDKVF---R 315
Db 342 ymgstveeghtkwvnyugnpvkrpprgaeidseaaanaaytslnlprkdtlqfvvhdprmd 401
Oy 316 ---TS---S-LK-G-YHSSLPRPRGKCL-PDQQPIR-TELFQVADBNRP-V-AQV-VE 360
Db 402 dsvrldnprllkdkvnylqvlvdrtdqaldgtvudvmfvsdtdgalhkal-s-lehav 458
Oy 361 -PMGRPLKT-P-LFHSXYHKQAVAHVQMASHGETHVLYLTDDRGITHHKVPEDCEHSF 417
Db 459 hi-leeqltqfderpqtlslsskgnrtvyagsnsgvvgarplafcgkhtvcedcydar 516
Oy 418 AFNFMELQPRRAAALQTMLSDAER-RKI-YVSSQWEEVSQVRLDCEVYGGCGGCLMSR 475
Db 517 dpycawspptecatvahqtcsparglligemsgdaavcpd-ks-kgsyrgfhfkngtucal 574
Oy 476 DPYCGMDQ--GRGISIYSBSRVLQSTINP-APRHKECPNPRDKAPLQKVSILANRSXYL 532
Db 575 kcsqknlavrfvfkfgnyvlkaespkyglmgtrknlillf-nlsegdsyvgylcsee 628
Oy 533 SCPEHSHNAVYSRHKENVEQCEP--GQASNCILFENTLTAOYGHYFCEAOE 585

```

RESULT 7
ID W51313 standard; Protein: 776 AA.
AC W51313;
DT 08-SEP-1998 (first entry)
DE Rat semaphorin W.
KW Rat; semaphorin W; nerve extension inhibitor; anti allergic; anticancer
KW immunosuppressant; gene therapy; diagnosis; research reagent.
OS Rattus norvegicus.
PN W09815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-PSDB: V07279, V07280.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Claim 1: Page 60-64; 90pp; Japanese.
CC The present sequence represents rat semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC anti allergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 776 AA;

	Query Match	12.0%	Score 510:	Df 32:	Length 776:
	Best Local Similarity	27.2%:	Pred. No.	3.0E-36:	
	Matches 151:	Conservative 136:	Mismatches 215:	Indels 53:	Gaps 37:
Dd	66 htnysalivdpashtcltyvgardisfgdlcflpfsgsrpridmwpvethrqncrkxkqke	124			
y	: : : : : :	:	:	:	:
92	OTPEHTVLFHFHPGSSWVGGRKRVY-L-FEPFGKNASV-RITNISTQSGSC--DKR-	92			


```
QY 148 VLFEGDEVYSTRKOEYNGKIPRFRIRGESELYTSDTVMONPOFIKATIVHODAYDDK 207
Db 218 lffffretaveyincgaiky-srvarvckhdkg-phqggdrwtstfiksrlncsvpgdyp 275
QY 208 IYFFREDNP-D-KNEAPLNVSRVAQLCRDQGGESLSVSKWNTFLKAMLVCSDAATNK 266
Db 276 fyfneiqstsdiegnvggq-vekliygvtftvnsiggsavcasfmsksilesfdgfk 334
QY 267 -FENRLQDV-F-LLPDP-SGQWRDRVYGVFSPNWN-Y--SAVCVYSL-G--D-IDKVFRT 316
Db 335 qetmmsnwlavpslkvppeprgqcvndsrtpdvsvnfvkshtlmdeavpafft-rpili 393
QY 317 -SSLKG-YHS--SL--PNPRFGKLPDOQPIPTFTFQVADRHPEVAQRVEPMGLKPLF 370
Db 394 rislgyrtfkivdgqvtgdkaydlfigtdgdkvikalnassfssdtdvdsvveel 453
QY 371 H-S-KYHYQKAV-HRMOASGETFHVLYLTDRGTTHKVVPEG--EOHSHFAFNIME-I 424
Db 454 qvlppgvpnklyvvrmdgddsklvvvsddellaiklhrsgsdktncrecvslqdpvca 513
QY 425 QPFRRAAAIQTW---SLDAERRKLYVSSQVEVSQVPLDLCEVYG-GGCHGCLMSRDPYCG 480
Db 514 wdnvelkctavgsdpds 530
QY 481 WDO-G-RCISYISSESS 495

RESULT 10
ID W51314 standard; Protein; 587 AA.
AC W51314;
DT 08-SEP-1998 (first entry)
DE Human semaphorin W.
KW Human; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PN W09815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMO ) SUMITOMO PHARM CO LTD.
PI Kkuchi K, Kimura T;
DR WPI; 98-261015/23.
DR N-PSDB; V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4: Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 10.1%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 2,10e-28;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedddelyffttstafdsyerikyprvarvcagdlgrktliq-grwtftlkadllcp 115
QY 201 DOAYDDKIYFFREDNPDKNEAPLNVSRVAQLCRDQGGESLSVSKWNTFLKAMLVCS 260
Db 116 gpehgrassvlqvavirpelga-gtptfygfsqqegatisavcafrpeditvngp 174
QY 261 DAAATKNFNRLQDVFLLPDPSPGQWRDRVYGVFSPNWN---YSACVYSLGIDKVFRTS 317
Db 175 freikhdcnrglpvvdvpprggecltnnmklrhfgsslsldrvltfirdhplndrp 234
QY 318 --SLK-GYHSSLP---N-----PRFGKCLPDQOPIP--TETFOVADRPE-V-AQR-VE-P 361
Db 235 vfpadghplvttdtaylrvrvahrvtlsqkeydvlylgtedghlhravrigaq-ls-vl 292
```

```
QY 362 MGPKLT-PLFHSK-YHYQKVAVHRMQASHGETFHVLYLTDRGTTHKVVPEGQEHSAF 419
Db 293 edial--fpepqvemkl-yhsw-llvgrtervtvnttcnrglqs-csecillaqdpvc 347
QY 420 NIMEIQFRRAAAIQTMSLDAERRKLYVSSQVEVSQVPLDLCEVYGCGCHGCLMSRDPYC 479
Db 348 awsfriqdecvahgehrqlvgqiesadvslcpkegprvfvfvpvataahvlpccps 407
QY 480 GWD-Q-Q-RCISYISSESSVLSQINPAEPHKECPNRPDKAPLQKVLAPNSRYLSCPME 537
Db 408 sawascvwhqpsgv 421
QY 538 SRHATYSWRHKNV 551

RESULT 11
ID W64221 standard; Protein; 974 AA.
AC W64221;
DT 06-OCT-1998 (first entry)
DE Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN W09827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GEMV ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 98-362774/31.
DR N-PSDB; V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17j; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 9.2%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 8,31e-25;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechfnfklknddalfvcctnafnpscrnykmdtlepfdefsgmarcypdakhn 172
QY 91 KRDCENTYIT-LLERRSEGLLACCAHNAHPSCWNLVNGTVVPLG-EMRYA--PSPSPENS 146
Db 173 valfadgklysatvt-dflaidaviyrl-gep--clrtvkhdkswklyefvqavdyg 228
QY 147 LVLFEGDEVYSTIRKOEYNG-KIPRFRIRGESELYTSDTVMONPOFIKATIVHODQAYD 205
Db 229 dylyffireaveyntmgkvvfrvraqvckndmgsgsrvlekqwtstfklrncsvpgds 288
QY 206 DKYIYFFREDNPDKNEAPLNVSRVAQLCRDQGGESLSVSKWNTFLKAMLVCSDAATN 265
Db 289 hfyfnilqavtdviring--rdv-vlatfstpynspgsavcaymdldiasvfgfrkeq 345
QY 266 K-NFNRLQDVFLLPDPSPGQWRDRVYGVFSPNWN-Y--SACVYSLGIDKVF--RTSSL 319
```

Db 346 kspdstwtpvderpkprgpcgagsssleryatsnefpddtlnfikthpImdeavpsif 405
 QY 320 KGYHSS---LNPGRKCLPD--QOPIPTETFOVADHRPEVA-Q--RVEP-M--G-P-L- 365
 Db 406 nrpflrtmryrltkiavdtaagpy-qhntvflgsekgaillkflarignsgfndslf 464
 QY 366 KTLPLF-HS--KYHYQKVAHVHRMQASHGETFHVLYLTDRGTIHKVVEP-GBQE--HSFAF 419
 Db 465 leemsvynsekcsydvgedkrimgmqldrasslyvafstcvikvpjgrcerhk-ckkt 523
 QY 420 -NIMEI-OPFFRA-AAQO---TMSLDAER--RKLYVSQWEVSQVPLDLCEVYGGCH-G 470
 Db 524 ciasrdpvcgwikeggacshlpsnrifecdiern 560
 QY 471 CLMSRDYCGW--DQGRCSISYSSERSVL-QSINPAE 504

RESULT 12
 ID R71384 standard; Protein; 712 AA.
 AC R71384;
 DT 21-NOV-1995 (first entry)
 DE Tribolium semaphorin I protein.
 KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
 KW variola major virus; smallpox; semaphorin receptor binding activity;
 KW modulation; nerve cell growth; immune response; viral pathogenesis;
 KW neurological disease; neuro-regeneration; oncological infection.
 OS Tribolium sp.
 PN W09507706-A.
 PD 23-MAR-1995.
 PF 13-SEP-1994; U10151.
 PR 13-SEP-1993; US-121713.
 PA (REGC) UNIV CALIFORNIA.
 PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
 PI O'Connor T;
 DR WPI; 95-131177/17.
 DR N-PSDB; Q87446.
 PT New class of semaphorin peptide(s) and polypeptide(s) - are
 PT potent modulators of nerve cell growth and regeneration
 PS Example 2; Page 85-88; 101pp; English.
 CC The sequence of the beetle Tribolium semaphorin I protein. The gene was
 CC isolated by PCR using Tribolium genomic DNA. The proteins encoded by the
 CC grasshopper semaphorin I (Q87441), human semaphorin III (Q87442),
 CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
 CC (Q87444-5), Tribolium semaphorin I or variola major (smallpox) virus
 CC semaphorin IV (Q87447) genes were used to generate a series of peptides
 CC (R70370-R70418), which retain semaphorin receptor binding activity. The
 CC semaphorin derived or semaphorin receptor derived peptides are potent
 CC modulators of nerve cell growth, immune responsiveness and viral
 CC pathogenesis. They can be used in diagnosis and treatment of neurological
 CC disease and neuro-regeneration, immune modulation and diagnosis and
 CC treatment of viral and oncological infection and diseases.
 SQ Sequence 712 AA;

Query Match 8.9%; Score 378; DB 13; Length 712;
 Best Local Similarity 27.0%; Pred. No. 1.12e-23;
 Matches 140; Conservative 128; Mismatches 194; Indels 56; Gaps 43;

Db 48 hflvng-detsilvggrnrvynlsifdlsrkggridwps-sdahqqlcllkgktddc 105
 QY 43 HTVLFHEPGSSVWVGGRKVY-L--PDFPEGKNASVTVNIGSTKGS-C-L-DKRD--C 94

Db 106 qnyirilyspepklvlgctnsykpqrtayafkegkylvekeviglcpynpnhnstsv 165
 QY 95 ENYITLLERRSEG--LLACGTNARHPSCWNLV--NGT-VVPLG-EMRGYAPDPDENSEL-V 148

Db 166 syng-qlfatsv-adfsggdpllyreqrtel-sdlqlnapfnvs--v----aydyi 216
 QY 149 LFEDEVYSTIRKQYNGKIPFRFRIRGESELYTSDTVNMQNPOPIKATIVHQOAYDDKI 208

Db 217 fffyreavemcngkvly-srvarvckdkgg-phgsrdwtksflkarlcnslpgeypf 274
 QY 209 YFFREDNPD-KNPEARLNVSRVAQLRGOGGESSLSVSKWNTFLKMLVCSDAATNK- 266

Db 275 yfelqstsdivegrynsddskliiygilttpeaynaigsgaicaayqmadillrvfegsfkhq 334
 QY 267 NFNLQDVFLLDPDSSGQWROTR--VYGVFSNPMN-Y--SAVCVYSLGDIDKVF---R-T 316
 Db 335 etinsnwpvpqnlvpeprpgqcvdrsiilpdknvnfikthslmed-vpalfg-kpvlvr 392
 QY 317 SSLKG-Y-H-S-SL-PNPRFGKCLPDQOPIPTETFOVADHRPEVAQRVEPMGPLKTLFLH 371
 Db 393 vslqyrftatvdqkvtkinnqyldvlyigtdgdkvilkavnipkrhakallyrkyrtsvh 452
 QY 372 -S-KYHYQKVAHV--RMOASHGETFHVLYLTDRGTIHKVVE-PGEQEHSEAFNIMEIQPF 427
 Db 453 phapavkqliagpygkvvvgkdeirlanlnhcas-ktckcdvelqdpchcawdaknl 511
 QY 428 RRAAIOTMSLDAERRKLYVSQWEVSQVPLDLCEVYGGCHGCLMSRDPYCGWD--QGR 485
 Db 512 cvsldtvsyrfliqdvrgddnk-cwspqtdkktvik 548
 QY 486 CISIYS-SE-RSVLQISINPAEPHKECNPDKRAPLOK 521

RESULT 13
 ID R71382 standard; Protein; 650 AA.
 AC R71382;
 DT 21-NOV-1995 (first entry)
 DE Drosophila semaphorin I protein.
 KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
 KW variola major virus; smallpox; semaphorin receptor binding activity;
 KW modulation; nerve cell growth; immune response; viral pathogenesis;
 KW neurological disease; neuro-regeneration; oncological infection.
 OS Drosophila sp.
 PN W09507706-A.
 PD 23-MAR-1995.
 PF 13-SEP-1994; U10151.
 PR 13-SEP-1993; US-121713.
 PA (REGC) UNIV CALIFORNIA.
 PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
 PI O'Connor T;
 DR WPI; 95-131177/17.
 DR N-PSDB; Q87444.
 PT New class of semaphorin peptide(s) and polypeptide(s) - are
 PT potent modulators of nerve cell growth and regeneration
 PS Example 2; Page 74-77; 101pp; English.
 CC The sequence of the Drosophila semaphorin I protein. The gene was
 CC isolated by PCR using primers based on sequence homology between the
 CC complete grasshopper and partial Tribolium and Manduca (moth) semaphorin
 CC gene sequences. The products gave two different sequences, each of which
 CC was used to obtain its respective complete sequence: semaphorin I and II
 CC (Q87445). The proteins encoded by the grasshopper semaphorin I (Q87441),
 CC human semaphorin III (Q87442), vaccinia virus semaphorin IV, Drosophila
 CC semaphorin I and II, Tribolium semaphorin I (Q87446) or variola major
 CC (smallpox) virus semaphorin IV (Q87447) genes were used to generate a
 CC series of peptides (R70370-R70418), which retain semaphorin receptor
 CC binding activity. The semaphorin derived or semaphorin receptor derived
 CC peptides are potent modulators of nerve cell growth, immune
 CC responsiveness and viral pathogenesis. They can be used in diagnosis and
 CC treatment of neurological disease and neuro-regeneration, immune
 CC modulation and diagnosis and treatment of viral and oncological infection
 CC and diseases.
 SQ Sequence 650 AA;

Query Match 8.7%; Score 367; DB 13; Length 650;
 Best Local Similarity 28.4%; Pred. No. 1.21e-22;
 Matches 124; Conservative 104; Mismatches 158; Indels 51; Gaps 38;

Db 3 dcqnyrimvpsgrlfcvgtstfrpmctyilsdntyleatknagvcydyprhnst 62
 QY 93 DCENYITLLERRSEG--LLACGTNARHPSCWNLV-LVNGTVVPLGEMR-GYA--PESPENSEL 147

Db 63 svladnelysgtv-adfsggdpllyreqlteqdydls-lnapnfv-ssft-qgd-f--- 114
 QY 148 VLFEDEVYSTIRKQYNGKIPFRFRIRGESELYTSDTVNMQNPOPIKATIVHQOAYDDK 207

QY 424 -IQPF---R----RAA--A--IQTMSILDAERRKLYVSSQWEVSQVPLDLCEVYGGCH-G 470
Db 529 clasqdpqgwhssrgcvcdirgsggtddvqagngesmehgdcqd 572
QY 471 CLMSRDPYCGWQGR-CISYISSE-RSVLQSIINPAE-PHKECPN 511

Search completed: Thu Jul 8 19:25:44 1999
Job time : 42 secs.

THIS PAGE BLANK (USPTO)

WQESRH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 19:22:36 1999; MasPar time 25.40 Seconds
Tabular output not generated. 895.992 Million cell updates/sec

Title: >US-09-041-236-2
Description: (39-606) from US09041236.pep (34 of 45)
Perfect Score: 4236
Sequence: 1 QTEPHVLFHEPGSSVWVG.....SYFERAQMOLLPEDGIMAE 568

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 48.996; Variance 86.541; scale 0.566

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1825	43.1	653	2	semaphorin homolog A3	0.00e+00
2	737	17.4	748	2	semaphorin A - mouse	2.04e-134
3	725	17.1	751	2	semaphorin E - mouse	1.18e-131
4	716	16.9	749	2	semaphorin V - human	1.38e-129
5	688	16.2	753	2	semaphorin III family	3.64e-123
6	681	16.1	772	2	collapsin - chicken	1.46e-121
7	669	15.8	772	2	semaphorin D - mouse	8.11e-119
8	667	15.7	666	2	semaphorin III - mouse	2.32e-118
9	661	15.6	771	2	semaphorin C - mouse	5.45e-117
10	618	14.6	782	2	semaphorin III precursor	3.39e-107
11	522	12.3	834	2	M-sema F protein prec	1.52e-85
12	480	11.3	760	2	semaphorin B - mouse	3.45e-76
13	470	11.1	403	2	A39R protein - vaccin	5.66e-74
14	463	10.9	441	2	hypothetical protein	2.00e-72
15	447	10.6	730	2	fasciclin IV precursor	6.73e-69
16	448	10.6	1074	2	semaphorin F precursor	4.06e-69
17	399	9.2	711	2	semaphorin I precursor	3.04e-56
18	367	8.7	656	2	semaphorin I - fruit	1.66e-51
19	362	8.5	724	2	semaphorin II precursor	1.95e-50
20	293	6.9	295	2	Sa1L9R protein - vacc	6.55e-36
21	196	4.6	1884	2	plexin 2 precursor -	1.10e-16
22	170	4.0	1872	2	plexin 3 precursor -	6.32e-12
23	168	4.0	1894	2	plexin 1 precursor -	1.44e-11

24 171 4.0 1905 2 I51553 Plaxin - African claw 4.19e-12
25 157 3.7 122 2 J31845 14R protein - variola 1.22e-09
26 157 3.7 122 2 H36852 A43R protein - variola 1.22e-09
27 155 3.7 142 2 QJ1776 Sa1R1R protein - vacc 2.71e-09
28 120 2.8 775 2 E70320 polynucleotide nu 1.32e-03
29 105 2.5 227 2 S09922 hypothetical protein 1.99e-01
30 105 2.5 235 1 Q3BEC9 HXLF4 protein - human 1.99e-01
31 102 2.4 275 2 A34866 T-cell surface protei 5.12e-01
32 103 2.4 406 2 B69064 conserved hypothetica 3.74e-01
33 99 2.3 275 2 S08464 T-cell alloantigen R 1.29e+00
34 99 2.3 460 2 S35772 translation elongatio 1.29e+00
35 99 2.3 563 2 S78224 virulence-associated 1.29e+00
36 99 2.3 591 2 S26565 virulence-associated 1.29e+00
37 99 2.3 591 2 S09498 virulence-associated 1.29e+00
38 99 2.3 591 2 S22664 virulence-associated 1.29e+00
39 99 2.3 593 2 S15215 virulence-associated 1.29e+00
40 99 2.3 597 2 A46050 thyroid/steroid recep 1.29e+00
41 98 2.3 641 2 S32017 flagellum-associated 1.74e+00
42 98 2.3 970 2 C57282 ankyrin-related prote 1.74e+00
43 98 2.3 979 2 B57282 ankyrin-related prote 1.74e+00
44 99 2.3 1375 2 JC5148 hepatocyte growth fac 1.29e+00
45 98 2.3 1786 2 A57282 ankyrin-related prote 1.74e+00

ALIGNMENTS

RESULT 1
ENTRY #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
T03102
T03102 #type complete
T14840
Ensser, A.; Pflanz, R.; Fleckenstein, B.
J. Virol. (1997) 71:6517-6525
Primary structure of the alcelaphine herpesvirus 1 genome.
T03102
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-653
#label ENS
#cross-references EMBL:AF005370; NID:g2337967; PID:g2337970
#length 653 #molecular-weight 73645 #checksum 5501

Query Match 43.1%; Score 1825; DB 2: Length 653;
Best Local Similarity 47.5%; Pred. No. 0.00e+00;
Matches 262; Conservative 95; Mismatches 184; Indels 11; Gaps 11;

Db 95 EPHTVLFHSLNSSDYVGGNNTIYLFDFAHSSNASTALINITSTHTHRLSSTCENFITL 154
Qy 41 EPHTVLFHEPGSSVWVGGRGVYLFDFPEGNASVTVNIGSTKSGCLDRDCENYITL 100
Db 155 LHNTQDGLLAGTNSQKPSCW-LINNLTQFLGPKGLAPSPSSGNLVLPQNDTYSTI 213
Qy 101 LERRSEGGLACGTNARHPSCNVLNVTVP-LGEMRGVAPESPENSLVLFEGDEVYSTI 159
Db 214 NLYKSLGSG-HKFRRIAGOVELYTSDTAHRRPQFOVATVHKNESYDDKIYFFFEENSHS 272
Qy 160 RK-QEYNGKIPFRRIIRGESELYTSDTVNQNPQFKALIVHQDQYDDKIYFFREDNPD 218
Db 273 DFQKQPTVPRVGVQSSDQGGESSLSYKWTTFTLKARLACVYDVTGRYNELODFIWO 332
Qy 219 KNPEAPLVNSVAQLCRGDQGGESSLSYKWTTFTLKARLACVYDVTGRYNELODFIWO 278
Db 333 APENSWEETLYGLFLSPWNSAVCVFTVKDIDHVFYKSLKKNYHKHLLPTPRPGCMKNH 392
Qy 279 DPSCQWRDTRVGVFSPNPNVSAVCVSLGDDIKVFTSSLKGYHSSLPNPRGKCLPDQ 338
Db 393 QHVPTEFQVADRYPEVADPVYQKNNMFPPIQSYIYTKLLVYRVEYG-GVFWATIFVL 451
Qy 339 QPIPTETQVADRHPEVAQRVEPMGPKLTPFHSHYKHKQKAVVHRMQASHGHTFH-VLYL 397
Db 452 TTIKGTIIHYRYEDSNSTTALNILEINPFQKPAPIQNLIDNTNLKLYVNSEWSEVSEVP 511

```
QY 398 TTDGTHHKVVEGEQHSFAFNIMEIQPFERRAAAIQTMSLDAERRKLYVSSQWV 457
      || :||| | : : : ||:| | ||:| :| | || :||:| :||| :| |
Db 512 LDLCVYGNDCFCFMSRDPLCTWYNTC-S-EK-QRVSVETGGPANRTLSEMGDHYAP 568
      ||| ||| | :||| | | :||| | | :||| | | :||| | | :||| | |
QY 458 LDLCEVYGGCGGCMRDPCYGDQGRCSISYSSERSVQDSINPAEPH-KE-CPNPKPD 515
      ||| :| | | | | | | | | | | | | | | | | | | | | | |
Db 569 TVVKHVSIPLLSNYSLSCPAVSNHADYFWTKDGFTEKRCHVKTHKNCDCILLIANSTAT 628
      ||| :| | | | | | | | | | | | | | | | | | | | | | |
QY 516 KAPLOKVS LAPNSRYLSLCPMESRHATYSWRHKNVQSCPEGHSNCILFIENLTAQQ 575
      ||| :| | | | | | | | | | | | | | | | | | | | | | |
Db 629 NGTHVCNKKEDS 640
      | :| :| | | | | | | | | | | | | | | | | | | | |
QY 576 YGHYFCEAQEGS 587
      | :| :| | | | | | | | | | | | | | | | | | | | |

RESULT 2 I48748 #type complete
ENTRY semaphorin A - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
DATE 04-Sep-1998
ACCESSIONS I48744
REFERENCE I48744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
#family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48744
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-748 #label RES
#cross-references EMBL:X85990; NID:g854323; PID:g854324
GENETICS
#gene sema
#superfamily semaphorin
CLASSIFICATION #length 748 #molecular-weight 82894 #checksum 9017
SUMMARY

Query Match 17.4%; Score 737; DB 2; Length 748;
Best Local Similarity 33.7%; Pred. No. 2.04e-134;
Matches 168; Conservative 112; Mismatches 166; Indels 52; Gaps 36;

Db 112 ECMFVRLHAYNTHLLACTGAFHPTCALRWATAGGTHASGPEKLEDGKGTYPDP 171
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 93 DCENITILLERSG-LLACTNARHPSG-----WNLVNGT-V-V-P--LGEMRGYAFSP 142
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 172 RHRPSPVLGELYSGV-TADLMGRDFTIFSLGQNPSLRTEPHDSRWLNPEKPVFWI 230
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 143 DENSLVLFEGDEVYSTIRKQYNGK-IPRRIRGESELYT-S-DTV-MQNPOFIKAT-I 197
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 231 PESENPDCKIYFFERSAEVAAPAMGRMSVSRVQICRNDLGGQRL-VNKTWTFELKAR 289
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 198 VHQAQYDDKIYFFREDNPDKNPE-APLNVSRAQLCRGDOGGESSLSVSKWNTFLKAM 256
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 290 LVCSPVGEVGDTHDQLQDVFLLS-SR-DROTPLLYAVFSTSGVFGSVCVYSMDVR 347
      ||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 257 LVCS--DAATNKNRNLQDVFLLPDPGQWRDTRVYGVFSNPWN-Y--SVCVYSLGDD 311
      ||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 348 RAFGLPLPHKEGPHQWYSGRVYPYPRPGMCPKSTGTFTSSSTKDFDDVQIFGRNPLM 407
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 312 KVFRTS-SLK-G-----YHSSLNPRPGKCLPDQ-QPIP-TETF--QVAD--R-HPEV 355
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 408 YNPVLMGG-R-PLFLQVAGYTTQTAAADRVAAADGH-YDVLFIGTDVGTVLKVISVPK 464
      ||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 356 AQRVEPMGPKTPLF-H--SKYHQKVAVHRMQASHGETHFVLYLTDRGTIHKVVE-P- 410
      ||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 465 GRRPNSGLLEELQVPEDSAAITSMOISSKROOLYVASRAAQAIALHRTALGRACAE 524
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 411 GQEHSAFNIMEIQPFERRAAAIQTMSLDAERRKLYVSSQWV 470
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 525 GCLARDPYCWDGSACTRFQPTAKRRFRQDIRNGDSTLCSGDSSHSVLEKKVLGVES 584
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
```

```
QY 471 CLMSRDPYCGWDQGRCSISYSSERSVL--QSINPAEPHKECPNPKDPKAPLQKVS LAPNS 528
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 585 GSAFLECEPRSLQAHVOW 602
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 529 -RYILSCPMSRHATYSW 545
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

RESULT 3 I48748 #type complete
ENTRY semaphorin E - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
DATE 04-Sep-1998
ACCESSIONS I48748
REFERENCE I48744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
#family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48748
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-751 #label RES
#cross-references EMBL:X85994; NID:g854331; PID:g854332
GENETICS
#gene sema
#superfamily semaphorin
CLASSIFICATION #length 751 #molecular-weight 85259 #checksum 8961
SUMMARY

Query Match 17.1%; Score 725; DB 2; Length 751;
Best Local Similarity 33.0%; Pred. No. 1.18e-131;
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;

Db 167 SFNPNTVSVMINELFSGMYI-DFMGTDAAIFRSLTKRMQLRTDQHNKSLSEPMFVD 225
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 139 PFPSPDENSELVLFEGDEVYSTIRKQYNGK-IPRRIRGESELYTS--DIV-MQNPOFIK 194
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 226 AHVPGDTPNDAKYVFFFKERLTDDNNRSTKQIHSMIARICPNDTGGQRL-VNKTWTFEL 284
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 195 ATIVHQD-QAYDDKIYFFREDNPDKNPEAPLNVSRAQLCRGDOGGESSLSVSKWNTFL 253
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 285 KARLVCSVDDEDCPETHFEDEVELL-ETDNP-RITLVYGIETSSSVFKGSAVCVYHL 342
      ||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 254 KAMLVCS--DA-ATNKNRNLQDVFLLPDPGQWRDTRVYGVFSNPWN-Y--SVCVYSL 307
      ||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 343 SDIOTVFNCGFAHKEGNHOLISYQGRIPYPRPGTCPPGAFTPNMRITTKDFPDVVTIFR 402
      ||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 308 GDIDKVFRTS-SLK-G-----YHSSLNPRPGKCLPDQ-QP-I-PTETF--QVADRHP 353
      ||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 403 NHPLMYSISPPIRRPLIVRIGTDYKYTKIADRVNAADG-RYHVLFLGTDRGTQKVVV 461
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 354 EVAQRVEPMGPL-KTPLF-H--SKYHQKVAVHRMQASHGETHFVLYLTDRGTIHKVVE 409
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 462 LPTNSASAGEVLEELVEFKNHVPIITMETISSKKOOLYSSNEGVSQVSLHRCHEYGTAC 521
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 410 -PGEQHSFAFNIMEIQPFERRAAAIQTMSLDAERRKLYVSSQWV 468
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 522 ADCCLARDPYCWDGHSRFRYPTGKRSSRRQDVRIHGNPLTQCRGNLKAIRNAEIVQY 581
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 469 HGCLMSRDPYCGWDQGRCSISYSS--RSVLQSIINPAEPHKECP--NPKDPKAPLQKVS 524
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 582 GVNNNSTEFCAPKSPQASIKLLOKDK 609
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 525 A-PNSRYILSCPMSRHATYSWR-HKEN 550
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

RESULT 4 G01856 #type complete
ENTRY semaphorin V - human
TITLE #formal_name Homo sapiens #common_name man
ORGANISM 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
DATE 04-Sep-1998
```



```
Qy 304 VYSLGDIKVERTS-SLK-G--YH--S--S-LPNPRGKCLPDQ-QPI-PTETF--QV-A 349
Db 402 FARSHAMPYNEVPEFINS-RPIMIKTVDVQFTQVIVDRVDAEDGQ-YDVMFIGTDIGTVL 459
Qy 350 -DR-HPEVAQVRPEMPGLKTPFLFSK--YHVKQVAVHRMQASHGETHFVLYLTIDRGTIIH 405
Db 460 KVSIPKETWHELEEVLEEMTVFREPTVISAMKISTKQOOLYIGTAGVSQLPLHRCDV 519
Qy 406 KVEE-PGEQHSFAFNIME-IQPFRAAAIQTMSLDAERKLYVSSQWESVQVPLDICEV 463
Db 520 YGKACAECLLARDPYCAWDGSSCSRYPTAKRTRRQDIRNGDPLTHCSDLOHHDNPSGQ 579
Qy 464 YGGCHGCLMSRDPYCGWDGRCISYSSE-RSVL-QSINPAEPHKECPN-PPDKAP-- 518
Db 580 TLEEKIYGVNSTFLECSPKSORAIVY-WOFQKQNDHDKVE 621
Qy 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQSC 556

RESULT 7
ENTRY semaphorin D - mouse
TITLE semaphorin D - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48747 #type complete
REFERENCE I48747 semaphorin D - mouse
#authors #formal_name Mus musculus #common_name house mouse
#journal Murine semaphorin D/collapsin is a member of a diverse gene
#title family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48747
GENETICS
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues 1-772 #label RES
#cross-references EMBL:X85993; NID:g854329; PID:g854330
#gene semD
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 772 #molecular_weight 88710 #checksum 1776

Query Match 15.8%; Score 669; DB 2; Length 772;
Best Local Similarity 30.2%; Pred. No. 8.11e-119;
Matches 170; Conservative 143; Mismatches 182; Indels 67; Gaps 51;

Db 58 HTFLDE-ERSRLVYGAKDHIFSNLVNIXDFQKIVMPVSY-TRRDECKWAGKDILKECA 115
Qy 43 HTVLFHEPGSSVWVGKGYLDFPPEGKN-AS-VRTVNIGSTKGSC--LDK---RDCE 95
Db 116 NFIKVEAYNOTHLYACGTGAFHPICTYIEVGHHPEDNIEFLQDSHFENGRGSPYDPKL 175
Qy 96 NYITLLERSEG-LLACGTNARHPSCNWL-V--N--GTVVPL--G--EM-RGYAPSPDE 144
Db 176 LTASLLIDGELYSGTA-ADEFMRDFAIFRTLGDHHPHPIRTEQHDNRWLNDPRFISAHLPE 234
Qy 145 NSLVLFEGDEVYSIRKQYNGK-IPFRIRIGESELVTS--DTV-MQNQFIKATIVHQ 200
Db 235 SDNPEDKVVYFFFRENAIGEHSGKATHARIGQICKNDFGGHRSL-VNKKWTFELKALIC 293
Qy 201 -DQAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKNTLKLMLVC 259
Db 294 SVCPNGIDTHFDELOQVFLM-NSKDP-KNPIYGVVFTTSSNIFKGSVMSDVRV 351
Qy 260 S--DA-ATNKNFNRLQDVFLLPDPSGQWRDTRVYGVFSNPNW-Y--SAVCVYSLGDIKV 313
Db 352 FLGPAHRDGPYQWVYQGRVYPRPGTCPSKTFGDFDSTKDLDPDVIITFGRSHPAMYN 411
Qy 314 FRTS-SLK-G--YH--S--S-LPNPRGKCLPDQ-QPI-PTETF--QV-A-DR-HPEVAQ 357
Db 412 PVFPINN-RPIMIKTVDVNYQFTQVIVDRVDAEDGQ-YDVMFIGTDIGTVLVKVPKRW 469
Qy 358 RVPEMPGLKTPFLFS--KYHYQKAVHRMQASHGETHFVLYLTIDRGTIIHKVVE-PGEQE 414
```

```
Db 470 HDLEEVLLEMTVFRPTTISAMELSTKQOOLYIGTAGVQAQLPHRCDIYKACAECCCL 529
Qy 415 HSFAFNIME-IQPFRAAAIQTMSLDAERKLYVSSQWESVQVPLDICEVYGGGCHGCLM 473
Db 530 ARDPYCAWDGSSCSRYPTAKRTRRQDIRNGDPLTHCSDLEHDHNNHGPSLEERIYGV 589
Qy 474 SRDPYCGWDGRCISYSSE-RSVL-QSINPAEPHKECPN-PPDK--AP-LQ-KVSLA- 525
Db 590 ENSSTFLECSPKSORALVY-WQ 610
Qy 526 PNSRYLSC-PMESRHATYSWR 546

RESULT 8
ENTRY semaphorin III - mouse (fragment)
TITLE semaphorin III - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I58169 #type fragment
REFERENCE I58169 semaphorin III - mouse (fragment)
#authors #formal_name Mus musculus #common_name house mouse
#journal Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;
#title Tessler-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#cross-references MUID:95267432
#accession I58169
GENETICS
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues 1-666 #label RES
#cross-references GB:L40484; NID:g703189; PID:g703190
#gene Semaphorin III
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 666 #checksum 9654

Query Match 15.7%; Score 667; DB 2; Length 666;
Best Local Similarity 30.1%; Pred. No. 2.32e-118;
Matches 150; Conservative 130; Mismatches 168; Indels 50; Gaps 39;

Db 60 RGKSPYDPKLLTASLLIDGELYSGTA-ANFMGRDFAIFRTLGHHPHPIRTEQHDNRWLNDP 118
Qy 135 RGVPSPDENSIVLEGEDEVYSTIRKQYNGK-IPFRIRIGESELVTS--DTV-MQNP 190
Db 119 RFTSAHLIPESNPEDKVVYFFFRENAIDGESHGKATHARIGQICKNDFGGHRSL-VNKKW 177
Qy 191 QFIKATIVHQ-DQAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKW 249
Db 178 TTFELKALICVPGPNCIDTHFDELOQVFLM-NSKDP-KNPIYGVVFTTSSNIFKGSVAVC 235
Qy 250 NTFELKALVCS--DA-ATNKNFNRLQDVFLLPDPSGQWRDTRVYGVFSNPNW-Y--SAVC 303
Db 236 MYSMSDVRVRLVLLGPAHRDGPYQWVYQGRVYPRPGTCPSKTFGDFDSTKDLDPDVIIT 295
Qy 304 VYSLGDIKV----F--RTSS-LK--GYHSSLPNPRGKCLPDQ-QPI-PTETF--QV-A 349
Db 296 FARSHAMPYNEVPEPINN-RPIMIKTVDVNYQFTQVIVDRVDAEDGQ-YDVMFIGTDIGTVL 353
Qy 350 -DR-HPEVAQVRPEMPGLKTPFLFS--KYHYQKAVHRMQASHGETHFVLYLTIDRGTIIH 405
Db 354 KVSVPKETHWHDLEEVLEEMTVFREPTTISAMELSTKQOOLYIGTAGVQAQLPLHRCDI 413
Qy 406 KVEE-PGEQHSFAFNIME-IQPFRAAAIQTMSLDAERKLYVSSQWESVQVPLDICEV 463
Db 414 YGKACAECLLARDPYCAWDGSSCSRYPTAKRTRRQDIRNGDPLTHCSDLOHHDNHPG 473
Qy 464 YGGCHGCLMSRDPYCGWDGRCISYSSE-RSVL-QSINPAEPHKECPN-PPDK--AP 518
Db 474 SLEERIYGVNSTFLECSPKSORALVYQWFORNEDRKEIKMGDHIIRTEQGLLLRS 533
Qy 519 -LQ-KVSLA-PNSRYLSC-PMESRHATY-SW-RHKNVEQSCPEG-HQ-SPNCILFIEN 570
```

```

Db 534 LOKKDSGNLYCHAVEHGF 551
      :|:|:| | |::
Qy 571 LTAQQYGHVFCEAQEGSY 588

RESULT 9
ENTRY semaphorin iii precursor - human
TITLE #formal_name Homo sapiens #common_name man
ORGANISM 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
DATE 04-Sep-1998

ACCESSIONS D49423 #type complete
REFERENCE D49423
#authors Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
#journal Cell (1993) 75:1389-1399
#title The Semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules.
#accession D49423
##status preliminary; nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-771 ##label KOL
GENETICS ##cross-references GB:L26081; NID:g799328; PID:g436560
#gene GDB:SEMA1
#cross-references GDB:283448
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 771 #molecular_weight 88889 #checksum 6249

Query Match 15.6%; Score 661; DB 2; Length 771;
Best Local Similarity 30.3%; Pred. No. 5.45e-117;
Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;

Db 166 RGKSPYPKLLTASLLDGLYSCTA-ADFMGRDAFAIRTLGHHPIRTQHDLSRWLNDP 224
      ||:|:| | |::
Qy 135 RGYAPFSDENSLVFEDEGVYSIRKQYNGK-IPFRIRGESELYTS--DTV-MQNPN 190
      ||:|:| | |::
Db 225 KFISAHLESNPEDDKYVFFFNENAIDEHSKATHARIGQICKNDFGGHRS-LVNKW 283
      ||:|:| | |::
Qy 191 QFIKATIVHQ-DQAYDDKIYYFFREDNPKNPEAPLVNSRVAQLCRGDQGESSLSYSKW 249
      ||:|:| | |::
Db 284 TTEFLKALICVPKPGINDITHEDELQDVFLM-NFKDP-KNPVVYGVFTTSSNIFKGSVC 341
      ||:|:| | |::
Qy 250 NTFLKMLVCS--DA-ATNKNFNLQDVFLLPDPGGQRDRTRVYGFSNPNW-Y--SAVC 303
      ||:|:| | |::
Db 342 MYMSDVRRVFLFGPYAHRDGPNIOWPVYQGRVPRPGTCPSKTGGFDSTKDLPDDVIT 401
      ||:|:| | |::
Qy 304 VYSGDIDKVERTS-SLK-G-YH-S-S-LPNRPKGCLPDQ-QPI-PTEF--QV-A 349
      ||:|:| | |::
Db 402 FARSHAMYNPVPMMNRPIVKTDVNYQTQIIVVDVDAEDGO-YDVMFIGTDVGTVLK 460
      ||:|:| | |::
Qy 350 -DR-HPEVAQRVEPMGLKPTPL-FHSKYHYQKVAVHRMQASHGETFHVLVLTDRGTIHK 406
      ||:|:| | |::
Db 461 VVISPKETWDLLEEVLEEMTVFREPTAISAMELSTKQOOLYGTAGVLAQLPLHRCDIY 520
      ||:|:| | |::
Qy 407 VVE-PGOEHSFAFNIME-IQPFRRAAIQTMSLDABERKLIVSSQWEVSQVPLDLCEVY 464
      ||:|:| | |::
Db 521 GKAAECCLARDPCAMDGACSACSFYPTAKRRTRRQDIRNGDPLTHCSDLHNDHNHGHP 580
      ||:|:| | |::
Qy 465 GGCGHGLMSRDPCGWNDQRCICIYSSE-RSVL-QSINPAEPHKECPNPKPK-----AP 518
      ||:|:| | |::
Db 581 EERIYGVENSSTFLECPKSQRALVYWFQFORNEEKRIIRDVDHIIRTDOGLLRSLQ 640
      ||:|:| | |::
Qy 519 LQKVSALA-PNSRYVLS-C-PMESRHATY-SW-RHKENV-EQSCERPQH-Q-SPNCILFIENLT 572
      ||:|:| | |::
Db 641 QKDSGNLYCHAVEHGFQ 658
      :|:|:| | |::
Qy 573 AQQYGHVFCEAQEGSYFR 590

RESULT 10
ENTRY semaphorin C - mouse (fragment)
TITLE

```

FEATURE							
1-21	#domain signal sequence #status predicted #label SIGV						
22-834	#product M-sema F protein #status predicted #label MAT						
SUMMARY	#length 834 #molecular-weight 92556 #checksum 7189						
Query Match 12.3%; Score 522; DB 2; Length 834;							
Best Local Similarity 28.5%; Pred. No. 1.52e-85;							
Matches 142; Conservative 118; Mismatches 190; Indels 49; Gaps 36;							
Db	102 KGKS-NOTECFNFIREFLOPYNSSHLYVCGYAFQPKACTYINMLTFTLDRAEFEDGKGP	160					
Qy	85 KGSCLKRDCENTITLLER-RSEGLACGTNAHPSC-W-NLVNGTV--VPJGEHMGVAP	139					
Db	161 YDPAKGHTGLLDGELYSATLN-NFLCTEPVILRYMGTHTHSIKTEYLAFWLNPEHFVGSA	219					
Qy	140 FSPDENSLVLFEDEVYSIRKOEYNGKTPRRIRGESELITS--TV-MQNPFQIKAT	196					
Db	220 FVPESVGSTGDDDKIYFFSERAVEYCYSEOVVARVAVCKGDMGMGARTLQ-KKWTF	278					
Qy	197 IVHQD-QAY---BDKIYYFFREDNPDKNPEAPLNVSRAQLCRGDGGESSLSVKWTF	252					
Db	279 LKARLVCSAPDWKYVNOLKAVHTLGAS--WHNTTFVGVOARWMDLSAVCEYOLEQ	336					
Qy	253 LKAWLVCSDAATNNRNLRQDFELLPPDSQGWDRTRVYGVSNPW---NYSAVCVSYSGD	309					
Db	337 IQQVFEGPYKEYSEQAKWARYTDVPVSRPGSCINNHRDNGYTSSLELPDNTLNFTKK	396					
Qy	310 IDKVF---RT-S-SLK-G-YHSSLNPNRPCKCLPD-QOPIP-TETFOVADHPPE-VA-	356					
Db	397 HPLMEQVOKPRGLRLVVKNTNTHVADVPLVDGATYTVLFITGDMGLLKAVSLGP	456					
Qy	357 QR-VE-PMGP-LKTPLFHSKY-HYQKVAVHRMQASHGETFHVLYLTDRGTTHKVVEPGE	412					
Db	457 WIH--M-VBELVFODEP-VESLVLSQSKVLFAGRSQQLVQSLADCTKYRF-CVDCV	510					
Qy	413 QEHSFAFNIEIQPFRRAAIIQTMSLDERRKLYVSSQEVSOVPLDJCEYVGGCGCL	472					
Db	511 LARDPYCAWNVTSRCVATTSGRSGLVQHVANLDTSKMCHYOYGIKKVRSIPKNIIVVS	570					
Qy	473 MSRDPCYGWD-Q-GRCISYSSER-SVL-QSINPAEPHKECNPDPDKA-PLOK-VSLAP	526					
Db	571 GTDLVLPCHLSSNLAAHW	589					
Qy	527 NSRYILSCPMESTRATYSW	545					
RESULT 12							
ENTRY	I48745	#title complete					
TITLE	semaphorin B - mouse						
ORGANISM	#formal_name Mus musculus #common_name house mouse						
DATE	02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 28-Feb-1997						
ACCESSIONS I48745							
REFERENCE	I48744						
#authors	Puschel, A.W.; Adams, R.H.; Betz, H.						
#journal	Neuron (1995) 14:941-948						
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension						
#cross-references MUID:95267431							
#accession	I48745						
#status	preliminary; translated from GB/EMBL/DDBJ						
#molecule_type	mRNA						
#residues	1-760 #label RES						
##cross-references EMBL:X85991; NID:g854325; PID:g854326							
GENETICS							
#gene	sempB						
SUMMARY	#length 760 #molecular-weight 83458 #checksum 2188						
Query Match 11.3%; Score 480; DB 2; Length 760;							
Best Local Similarity 29.2%; Pred. No. 3.45e-76;							
Matches 157; Conservative 121; Mismatches 203; Indels 57; Gaps 41;							
Db	123 QCENFIRVLVSYNATHLYACGTAFAFPACTFIELQDSLPLILDVKVMGKGQSPLTLFT	182					

Db	302	EVIEKYNVLDDIIKPLS	318
Qy	347	QVADRRHPEVAQRVEPMG	363
RESULT	14		
ENTRY	S29921	#type complete	
TITLE	hypothetical protein 15 - vaccinia virus		
ORGANISM	#formal_name vaccinia virus		
DATE	20-Feb-1995	#sequence_revision	20-Feb-1995
	09-Sep-1997	#text_change	
ACCESSIONS	S29921		
REFERENCE	S29907		
#authors	Amegadzie, B.Y.		
#submission	submitted to the EMBL Data Library, January 1991		
#accession	S29921		
#status	preliminary		
#molecule_type	DNA		
#residues	1-441	#label AME	
#cross-references	EMBL:X57318; NID:g62239; PID:g62254		
SUMMARY	#length 441	#molecular-weight 50185	#checksum 6034
Query Match	10.9%;	Score 463;	DB 2; Length 441;
Best Local Similarity	33.9%;	Pred. No. 2.00e-72;	
Matches	87;	Conservative	57; Mismatches 96; Indels 17; Gaps 16;
Db	115	LVCSTNNPKCKW-IDGSDDPKHRGYAPYQNSKVTIIHNGC-VLSDINISK-EG-I	170
Qy	109	LACGTNARHPSNVLNGTVVPLCEMRGYAPFSPDENSFLVEGDEVYSTIRKOEYNGKI	168
Db	171	KRWRFPGCGYDLATADNVIKDG-LRGAFAVDKGTID-KVILFTDTIGSKR-I-VK	225
Qy	169	PRFRIRIGE-S-ELYTSDTVQNPQFIKATIVHQDQAYDDKIYFFREDNPDKNPEAPLN	226
Db	226	IPYHAQCLNDEGSPSSSHRNSTFLKVELEC-DID-GRSY-R-Q-IHSRTIKTD-ND	279
Qy	227	VSRYAQLCRDGGESSLSVSKWNTFLKMLVCSDAATNKNFNRLQDVFLLPDPGSGWRD	286
Db	280	TILVFFDPSYSKALCTYSMTTIKQSFSTSKLEGYTKQLPSPASGTCCLPAGKVPHTTF	339
Qy	287	TRVGVFSNPWNYSACVSYLGIDKDKVFTSSSLKGYHSSLPNPRPGKCLDQOPIPTET	346
Db	340	EVIEKYNVLDDIIKPLS	356
Qy	347	QVADRRHPEVAQRVEPMG	363

Search completed: Thu Jul 8 19:23:23 1999

Job time : 47 secs.

```

RESULT      15
ENTRY       JH0798      #type complete
TITLE       fasciclin IV precursor - American bird grasshopper
ORGANISM    #formal_name Schistocerca americana fcommon_name American
              bird grasshopper
DATE         30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
              09-Sep-1997
ACCESSIONS  JH0798
REFERENCE    JH0798
#authors     Kolodkin, A.; Matthes, D.J.; O'Connor, T.P.; Patel, N.H.;
              Admon, A.; Bentley, D.; Goodman, C.S.
#journal     Neuron (1992) 9:831-845
#title       Fasciclin IV: Sequence, expression, and function during
              growth cone guidance in the grasshopper embryo.
#accession   JH0798
              ##molecule_type mRNA
              ##residues      1-730 ##label KOL
              ##cross-references GB:L00709; NID:g160844; PID:g160845
              ##experimental_source embryo
COMMENT      This protein plays a role in growth cone guidance in the developing
              central nervous system.
              glycoprotein; transmembrane protein
KEYWORDS
FEATURE      1-22      #domain signal sequence #status predicted #label SIG\
23-730          #product fasciclin IV #status predicted #label MAT\
23-627          #domain extracellular #status predicted #label EXT\

```

THIS PAGE BLANK (USPTO)

WQREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:21:57 1999; Maspar time 17.91 Seconds
Tabular output not generated. 896.325 Million cell updates/sec

Title: >US-09-041-236-2
Description: (39-606) from US09041236.pap (34 of 45)
Perfect Score: 4236
Sequence: 1 QTEPHTVLFHPGSSVWVG.....SYFREAQHWQLLPEDGIMAE 568

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: swiss-prot37
1:swissprot

Statistics: Mean 50.069; Variance 75.814; scale 0.660

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	%	ID	Description	Pred. No.
1	470	11.1		1 VA39_VACCC	PROTEIN A39.	3.32e-86
2	463	10.9		1 VA39_VACCV	PROTEIN A39.	2.09e-84
3	163	3.8		1 SEX_HUMAN	TRANSMEMBRANE PROTEIN H	8.53e-13
4	105	2.5		1 US08_HCMVA	HYPOTHETICAL PROTEIN H	3.58e-02
5	102	2.4		1 NRT2_RAT	T-CELL ECTO-ADP-RIBOSY	1.04e-01
6	99	2.3		1 NRT1_RAT	T-CELL ECTO-ADP-RIBOSY	2.97e-01
7	99	2.3		1 EFLA_TRIE	ELONGATION FACTOR 1-AL	2.97e-01
8	99	2.3		1 VRP2_SALT	65 KD VIRULENCE PROTEI	2.97e-01
9	99	2.3		1 VRP2_SALCH	65 KD VIRULENCE PROTEI	2.97e-01
10	99	2.3		1 VRP2_SALEN	65 KD VIRULENCE PROTEI	2.97e-01
11	99	2.3		1 VRP2_SALDU	65 KD VIRULENCE PROTEI	2.97e-01
12	99	2.3		1 PMAJ_AJECA	PLASMA MEMBRANE ATPASE	2.97e-01
13	98	2.3		1 DNBI_HSV6U	MAJOR DNA-BINDING PROT	4.19e-01
14	95	2.2		1 HA21_HUMAN	H2A CLASS II HISTOCOMP	1.15e+00
15	94	2.2		1 RBSR_ECOLI	RIBOSE OPERON REPRESSO	1.61e+00
16	93	2.2		1 DP3E_MYCPN	DNA POLYMERASE III, BE	2.23e+00
17	92	2.2		1 CEMI_YEAST	3-OXOACYL-(ACYL-CARRIE	3.09e+00
18	92	2.2		1 EFLA_NEUCR	ELONGATION FACTOR 1-AL	3.09e+00
19	95	2.2		1 VL2_HPV5B	MINOR CAPSID PROTEIN L	1.15e+00
20	95	2.2		1 VL2_HPV05	MINOR CAPSID PROTEIN L	1.15e+00
21	95	2.2		1 AMOH_ARTGO	HISTAMINE OXIDASE (EC	1.15e+00
22	94	2.2		1 PMA2_YEAST	PLASMA MEMBRANE ATPASE	1.61e+00
23	94	2.2		1 POL_HTLV2	POL POLYPROTEIN [CONTA	1.61e+00

24	93	2.2	1029	1	END1_YEAST	VACUOLAR BIOGENESIS PR	2.23e+00
25	94	2.2	1069	1	ENTK_MOUSE	ENTEROPEPTIDASE (EC 3.	1.61e+00
26	94	2.2	1115	1	IRE1_YEAST	SERINE/THREONINE-PROTE	1.61e+00
27	92	2.2	1132	1	DNBI_HSV62	MAJOR DNA-BINDING PROT	3.09e+00
28	94	2.2	1200	1	DXB8_CAEEL	PUTATIVE PRE-MRNA SPLI	1.61e+00
29	92	2.2	1302	1	MDR5_DROME	MULTIDRUG RESISTANCE P	3.09e+00
30	95	2.2	1390	1	MET_HUMAN	HEPATOCYTE GROWTH FACT	1.15e+00
31	92	2.2	1663	1	CO3_MOUSE	COMPLEMENT C3 PRECURSO	3.09e+00
32	90	2.1	1119	1	RL19_MYCPN	50S RIBOSOMAL PROTEIN	5.84e+00
33	90	2.1	213	1	AMEX_BOVIN	AMELOGENIN, CLASS I PR	5.84e+00
34	90	2.1	254	1	HA22_HUMAN	H2A CLASS II HISTOCOMP	5.84e+00
35	91	2.1	348	1	DDL_ENTFA	D-ALANINE-D-ALANINE L	4.25e+00
36	91	2.1	356	1	VP39_NPVLD	MAJOR CAPSID PROTEIN.	4.25e+00
37	89	2.1	404	1	NIFS_ECOLI	NIFS PROTEIN HOMOLOG.	7.99e+00
38	89	2.1	437	1	RFBB_MYXXA	O-ANTIGEN EXPORT SYSTE	7.99e+00
39	91	2.1	498	1	ELAS_PSEAE	PSEUDOLYSIN PRECURSOR	4.25e+00
40	91	2.1	598	1	NOT_HUMAN	IMMEDIATE-EARLY RESPON	4.25e+00
41	89	2.1	598	1	RNR1_RAT	REGENERATING LIVER NUC	7.99e+00
42	89	2.1	649	1	RA32_SCHPO	DNA REPAIR PROTEIN RAD	7.99e+00
43	91	2.1	707	1	GCVK_HCMVA	GANCICLOVIR KINASE (EC	4.25e+00
44	91	2.1	1666	1	CO3_CAVPO	COMPLEMENT C3 PRECURSO	4.25e+00
45	90	2.1	1874	1	POLR_KYMVJ	RNA REPLICASE POLYPROT	5.84e+00

ALIGNMENTS

RESULT 1
ID VA39_VACCC STANDARD; PRT: 403 AA.
AC P21062;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE PROTEIN A39.
GN A39R.
OS VACCINIA VIRUS (STRAIN COPENHAGEN).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91021027.
RA GOBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RT "The complete DNA sequence of vaccinia virus.";
RL VIROLOGY 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA GOBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RL VIROLOGY 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M35027; G335517; -
DR PIR: E42521; E42521.
SQ SEQUENCE 403 AA; 45741 MW; FE0DB9AB CRC32;
Query Match 11.1%; Score 470; DB 1: Length 403;
Best Local Similarity 33.5%; Pred. No. 3.32e-86;
Matches 86; Conservative 59; Mismatches 95; Indels 17; Gaps 15;

Db 77 LVCGTNGNPKWC - IDGDDDPKHRGYPAYQNSKVTIISYN-ECVLSIDNISK-EG-I 132
QY 109 LACGTNARHPSCNVLVNGVTVPGLGEMRGVAPSPDENSILVLFEGDEVYSTIRKQEVNGKI 168
Db 133 KWRFRFDGCGVLDLTADNVIPKDG-LRGAFVKGDKTYD-KVYILFTDTIGSKR--I-VK 187
QY 169 PRFRIRGE-S-ELTSDTVMQNPQFIKATIVHQDAQYDKKIYFFREDNPDKNPEAPLN 226

QY 131 LGE-MRGYAPFSPDENSILVFEDEVYSTRKOEYNGK-IPRFRIRGESE-LYTSDTVM 187
Db 252 ES 253
QY 188 QN 189

RESULT 6
ID NRT1_RAT STANDARD; PRT; 275 AA.
AC P17982;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 (T-CELL
NAD(P)(+)-L-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL MONO(ADP-
RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN
RT6.1)
GN ART2A OR RT6-A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-LEWIS A;
RX MEDLINE; 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell
alloantigen RT6.1."
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
[2]
RN MUTAGENESIS OF GLN-207.
RX MEDLINE; 96275529.
RA MAEHAMA T., HOSHINO S.-I., KATADA T.;
RT "Increase in ADP-ribosyltransferase activity of rat T lymphocyte
alloantigen RT6.1 by a single amino acid mutation."
RL FEBS LETT. 388:189-191(1996).
CC -!- FUNCTION: HAS NAD+ GLYCOHYDROLASE ACTIVITY AND EXTREMELY LOW ADP-
RIBOSYLTRANSFERASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -!- CATALYTIC ACTIVITY: NAD(+) + H2O = NICOTINAMIDE + ADP-RIBOSE.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -!- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X52082; G57168; -.
DR EMBL; M31138; G206804; -.
DR PIR; S08464; S08464.
DR PFAM; P501291; ART; 1.
DR PROSITE; P501291; ART; 1.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
KW T-CELL DIFFERENTIATION; GPI-ANCHOR.
FT SIGNAL 1 20
FT CHAIN 21 246
FT PROPEP 247 275
FT T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
FT HYDROPHOBIC, REMOVED DURING MATURATION
FT (BY SIMILARITY).
FT GPI-ANCHOR (BY SIMILARITY).
FT POTENTIAL.
FT Q->E: INCREASED ADP-RIBOSYLTRANSFERASE
FT ACTIVITY.
FT MUTAGEN 207 207
FT MUTAGEN 207 207
SQ SEQUENCE 275 AA; 31388 MW; 523B1A84 CRC32;

Query Match 2.3%; Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 2.97e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;
Db 194 LGVIKESF-PYDOEE-VLIPGEVYOKVTRQTGYNELFDSPPRKKSNYNCLYSSAGTR 251
QY 131 LGE-MRGYAPFSPDENSILVFEDEVYSTRKOEYNGK-IPRFRIRGESE-LYTSDTVM 187
Db 252 ES 253
QY 188 QN 189

RESULT 7
ID EF1A_TREIR STANDARD; PRT; 460 AA.
AC P34825; 1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEF1.
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-OM9414 / RUTC-30;
RX MEDLINE; 94124021.
RA NAKARI T., ALATALO E., PENTTILA M.;
RT "Isolation of Trichoderma resei genes highly expressed on glucose-
containing media: characterization of the tef1 gene encoding
translation elongation factor 1 alpha."
RL GENE 136:313-318(1993).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
EF-TU/EF-1A SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; Z23012; G312887; -.
DR PIR; S35772; S35772.
DR PROSITE; PS00301; EFACOR_GTP; 1.
DR PFAM; PF00009; GTP_EFTU; 1.
DR HSPM; P07157; IAIP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 49830 MW; B9ABAB2 CRC32;
Query Match 2.3%; Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 2.97e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;
Db 210 YGWEKETKAGFTG-KTLLEAIDSIEPPKR-PTDKPLRLPLQDV 252
QY 478 YCGWDQGRCSISYSERSVLOSINPAEPHKECPNPKPKAPLQKV 522

RESULT 8
ID VRP2_SALTY STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN MKAA.
OS SALMONELLA TYPHIMURIUM.
OC PLASMID 96 KB VIRULENCE PEX102.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TML R66;
RX MEDLINE; 90136009.
RA TAIRA S., RHEN M.;
RT "Identification and genetic analysis of mkaA -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth".
RL MICROB. PATHOG. 7:165-173(1989).
RN [2]
RP SEQUENCE OF 1'-10.
RX MEDLINE; 91244158.
RA TAIRA S., BAUMANN M., RIIKONEN P., SUKUPOLVI S., RHEN M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium".
RL FEMS MICROBIOL. LETT. 61:319-323(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15042; G47783; -
DR PIR; A54540; A54540.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 POLY-PRO.
SQ SEQUENCE 591 AA; 65440 MW; B69F25C3 CRC32;
Query Match 2.3%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 2.97e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDNSGILHLGTAARLSDPOAASHTAQW-LVEESVTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RCENYITLLRRSEGLACGTNARHPSCNVLNGTVPLGEMRGYAPFSPDENSILVLF 151
Db 220 GNEAGDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRIRGESELY 181
RESULT 9
ID VRP2_SALCH STANDARD; PRT; 591 AA.
AC P17450;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 55 KD VIRULENCE PROTEIN (PROTEIN M2, IN MBA REGION).
OS SALMONELLA CHOLERAE-SUIS (SALMONELLA ENTERICA).
OG PLASMID PKDSC50.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RF-1;
RX MEDLINE; 90245675.
RA MATSUI H.;
RT "Nucleotide sequences of genes encoding 32 kDa and 70 kDa
RT polypeptides in mba region of the virulence plasmid, pKDC50, of

RT Salmonella choleraesuis".
RL NUCLEIC ACIDS RES. 18:2181-2181(1990).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52035; G46898; -
DR PIR; S09498; S09498.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 POLY-PRO.
SQ SEQUENCE 591 AA; 65341 MW; BEA439A CRC32;
Query Match 2.3%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 2.97e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDNSGILHLGTAARLSDPOAASHTAQW-LVEESVTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RCENYITLLRRSEGLACGTNARHPSCNVLNGTVPLGEMRGYAPFSPDENSILVLF 151
Db 220 GNEAGDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRIRGESELY 181
RESULT 10
ID VRP2_SALEN STANDARD; PRT; 591 AA.
AC P55220;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN SPVB.
OS SALMONELLA ENTERITIDIS.
OG PLASMID PNL2001.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AL1190;
RX MEDLINE; 94362897.
RA SUZUKI S., KOMASE K., MATSUI H., ABE A., KAWAHARA K., TAMURA Y.,
RA KIJIMA M., DANBARA H., NAKAMURA M., SATO S.;
RT "Virulence region of plasmid PNL2001 of Salmonella enteritidis".
RL MICROBIOLOGY 140:1307-1318(1994).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14490; G517164; -
DR PLASMID; VIRULENCE.
FT DOMAIN 367 POLY-PRO.
SQ SEQUENCE 591 AA; 65350 MW; 9C5B2EE3 CRC32;
Query Match 2.3%; Score 99; DB 1; Length 591;

```
Best Local Similarity 23.3%; Pred. No. 2.97e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLCKTAARLSDPQAASHTAQW-LVESVTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RDCENYITLLRRSGGLACGTNARHPSCWNLVNGTVVPLGEMRGYAPFSPDENSILVFE 151
Db 220 GNEAGDRSAMRYSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRRIIRGESELY 181

RESULT 11
ID VRP2-SALDU STANDARD; PRT: 593 AA.
AC P24419;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN VSDC.
OS SALMONELLA DUBLIN.
OC PLASMID PSDL2.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91251759.
RA KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINEY D.;
RT "Molecular analysis of the virulence locus of the Salmonella dublin
plasmid PSDL2.";
RL MOL. MICROBIOL. 5:307-316(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X56727; G47839; -
DR PIR: S15215; S15215.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 375 POLY-PRO.
SQ SEQUENCE 593 AA; 65613 MW; A7CE57B5 CRC32;

Query Match 2.3%; Score 99; DB 1; Length 593;
Best Local Similarity 23.3%; Pred. No. 2.97e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLCKTAARLSDPQAASHTAQW-LVESVTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RDCENYITLLRRSGGLACGTNARHPSCWNLVNGTVVPLGEMRGYAPFSPDENSILVFE 151
Db 220 GNEAGDRSAMRYSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRRIIRGESELY 181

RESULT 12
ID PMAL-AJECA STANDARD; PRT: 916 AA.
AC Q07421;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).
GN PMAL.
OS AJELLOWYCES CAPSULATA (HISTOPLASMA CAPSULATUM).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;

Best Local Similarity 23.3%; Pred. No. 2.97e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLCKTAARLSDPQAASHTAQW-LVESVTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RDCENYITLLRRSGGLACGTNARHPSCWNLVNGTVVPLGEMRGYAPFSPDENSILVFE 151
Db 220 GNEAGDRSAMRYSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRRIIRGESELY 181

RESULT 13
ID DNBI-LHSV6U STANDARD; PRT: 1132 AA.
AC P52338;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MAJOR DNA-BINDING PROTEIN (MDBP).
GN UAL.
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; ROSEOLOVIRUS.
```

```
OC ONYGENALES; ONYGENACEAE; AJELLOWYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94124018.
RA SCHAFER M.P., DEAN G.E.;
RT "Cloning and sequence analysis of an H(+)-ATPase-encoding gene from
the human dimorphic pathogen Histoplasma capsulatum.";
RL GENE 136:295-300(1993).
CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING
EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKINIZATION MAY MEDIATE
GROWTH RESPONSES.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O -> ADP + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
(EI-E2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L07305; G409249; -
DR PROSITE: PS00154; ATPASE_E1_E2; 1.
DR PFAM: PF00122; EI-E2_ATPase; 1.
KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
KW ATP-BINDING.
FT DOMAIN 1 111 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 112 134 1 (POTENTIAL).
FT DOMAIN 135 136 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 137 155 2 (POTENTIAL).
FT DOMAIN 156 279 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 280 305 3 (POTENTIAL).
FT DOMAIN 306 320 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 321 350 4 (POTENTIAL).
FT DOMAIN 351 711 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 712 734 5 (POTENTIAL).
FT DOMAIN 735 754 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 755 776 6 (POTENTIAL).
FT DOMAIN 777 817 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 818 843 7 (POTENTIAL).
FT DOMAIN 844 847 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 848 874 8 (POTENTIAL).
FT DOMAIN 875 916 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 374 374 ASP/GLU-RICH (ACIDIC).
FT BINDING 470 470 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 916 AA; 98884 MW; 4A468A44 CRC32;

Query Match 2.3%; Score 99; DB 1; Length 916;
Best Local Similarity 37.2%; Pred. No. 2.97e-01;
Matches 16; Conservative 8; Mismatches 16; Indels 3; Gaps 3;

Db 762 LAVGTWIT-ITTMLVGSENGIIVONFORTHVPLFLEISLTENW 803
QY 244 LSVSKWNTFLKMLVCS-D-AATNKNFNLQDV-FLPDPSQOW 284

RESULT 13
ID DNBI-LHSV6U STANDARD; PRT: 1132 AA.
AC P52338;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MAJOR DNA-BINDING PROTEIN (MDBP).
GN UAL.
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; ROSEOLOVIRUS.
```



```

[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95266321.
RA COMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.:
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RL and genome evolution.";
RL VIROLOGY 209:29-51(1995).
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES' DNA-BINDING PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X92436; G1044871; -.
DR EMBL; X83413; G854020; -.
DR PFAM; PF00747; viral_DNA_bp; 1.
KW DNA-BINDING; DNA REPLICATION; ZINC-FINGER; NUCLEAR PROTEIN.
FT ZN.FING 459 475 C4-TYPE
SQ SEQUENCE 1132 AA; 127762 MW; DB699A02 CRC32;

Query Match 2.3%; Score 98; DB 1; Length 1132;
Best Local Similarity 28.2%; Pred. No. 4.19e-01;
Matches 24; Conservative 18; Mismatches 37; Indels 6; Gaps 6;

Db 435 WNLNRMVYNAGNATEI-YNHLVNCNANLCEFCGKC-C-QSCIGTAMRVGTRLPAP 491
Qy 297 WNYSAVCYSLGDDIKVFTTSKLGKHSLSLNPREGKLPDQOPIPTETQVADRHPVA 356

Db 492 KNVK-KEPLVMSMF-SRY-YAEVDI 513
Qy 357 QRVEPMGLKPLFLSKHYQKAV 381

RESULT 14
ID HA21 HUMAN STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84168117.
RA AUFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPES D.,
RA STROMINGER J.L.;
RT "Isotypic and allotypic variation of human class II
RT histocompatibility antigen alpha-chain genes.";
RL NATURE 308:327-333(1984).
RN [2]
RP SEQUENCE OF 40-255 FROM N.A.
RX MEDLINE; 83065171.
RA AUFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT "cDNA clone for the heavy chain of the human B cell alloantigen DC1:
RT sequence homology to the HLA-DR heavy chain.";
RL PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J00199; -. NOT_ANNOTATED_CDS.
CC
CC PIR; A02212; HLHUDC.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00993; MHC_II_alpha; 1.
DR HSSP; P01910; IIAK.
KW MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 23
FT CHAIN 24 255 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT DO(1) ALPHA CHAIN.
FT DOMAIN 24 110 EXTRACELLULAR ALPHA-1.
FT DOMAIN 111 204 EXTRACELLULAR ALPHA-2.
FT DOMAIN 205 217 CONNECTING PEPTIDE.
FT TRANSMEM 218 240
FT DOMAIN 241 255 CYTOPLASMIC TAIL.
FT DISULFID 133 189 BY SIMILARITY.
FT CARBOHYD 104 104 POTENTIAL.
FT CARBOHYD 144 144 POTENTIAL.
SQ SEQUENCE 255 AA; 28105 MW; 5C46DD59 CRC32;

Query Match 2.2%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1.15e+00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;

Db 25 DIVADSVAGLG-VNLYQSGPQSGYSHFEDGDEEYVDLERKTVQLPLRFR 78
Qy 122 NLVNGTVVPLGEMRGYAPFSPDENSILVLFEGDEV-YSTIRKQYNGKIPRRIR 175

RESULT 15
ID RBSR_ECOLI STANDARD; PRT; 329 AA.
AC P25551;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOSE OPERON REPRESSOR.
GN RBSR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6 AND 250-256.
RC STRAIN-K12;
RX MEDLINE; 93278299.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural and functional analyses of the repressor, RbsR, of the
RT ribose operon of Escherichia coli.";
RL PROTEIN SCI. 1:831-842(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 93315143.
RA BURLAND V.D., PLONKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL GENOMICS 16:551-561(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA IIDA A., TESHIBA S., MIZOBUCHI K.;
RT SUBMITTED (XXX-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SIMILARITY TO RIBOSE-BINDING PROTEINS.
RX MEDLINE; 93278300.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural homology between rbs repressor and ribose binding protein
RT implies functional similarity.";
RL PROTEIN SCI. 1:843-849(1992).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RESDACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING
```

```
CC THE RBS OPERON TRANSCRIPTIONAL START SITE. THE AFFINITY FOR THE
CC RBS OPERATOR IS REDUCED BY ADDITION OF RIBOSE, CONSISTENT WITH
CC RIBOSE BEING THE INDUCER OF THE OPERON.
CC -|- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13169; G147517; -.
DR EMBL; L10328; G290603; ALT_INIT.
DR EMBL; AE000452; G1790194; -.
DR EMBL; D10466; G471110; -.
DR PIR; A41828; A41828.
DR ECOGENE; EG10819; RBSR.
DR PROSITE; PS00356; HTH_LACI_FAMILY; 1.
DR PFAM; PF00356; lacI; 1.
DR PFAM; PF00532; Peripla_BP_like; 1.
DR HSP; P15039; lprv.
DR KW TRANSCRIPTION REGULATION; REPRESSOR; DNA-BINDING.
FT INIT_MET 0
FT DNA_BIND 3 22 H-T-H MOTIF (BY SIMILARITY).
FT CONFLICT 11 12 GV -> L (IN REF 1).
SQ SEQUENCE 329 AA; 36480 MW; 424E72A6 CRC32;

Query Match 2.2%; Score 94; DB 1; Length 329;
Best Local Similarity 34.6%; Pred. No. 1.61e+00;
Matches 18; Conservative 14; Mismatches 14; Indels 6; Gaps 6;

Db 109 TLMQKRVGGLLLC-TETHQPS-REIMQRYPTVP-TVMMDWAPFGDSD-LI 156
Oy 99 TLLERRSEGLLA-CGTNARHPSCNVLNG-TVVP LGEMRGYAPFSPDENSIV 148
```

Search completed: Thu Jul 8 19:22:19 1999
Job time : 22 secs.

Db	371	QPIPTTTFQVADRHPEVAQRVEPMGPKLTPLFHSKSYHQKVAVHRMQASHGETFHVLYLT	430
Qy	339	QPIPTTTFQVADRHPEVAQRVEPMGPKLTPLFHSKSYHQKVAVHRMQASHGETFHVLYLT	398
Db	431	TDRTGTHKKVVEPGEQEHGFAFNMIEIQPFERRAAAIQTMSLDAERRKLYVSSQWEVSQVPL	490
Qy	399	TDRTGTHKKVVEPGEQEHGFAFNMIEIQPFERRAAAIQTMSLDAERRKLYVSSQWEVSQVPL	458
Db	491	DLCEVYGGCHGLMSRDPYCGWDGRCISYISYSSERSVLQSIINPAEPHKECNPXPKDPKAP	550
Qy	459	DLCEVYGGCHGLMSRDPYCGWDGRCISYISYSSERSVLQSIINPAEPHKECNPXPKDPKAP	518
Db	551	LQKVS LAPNSRYLLSCPME SRHATYSWRHKNVQSCPEGHQSPNCILFIENLTAAQQYGH	610
Qy	519	LQKVS LAPNSRYLLSCPME SRHATYSWRHKNVQSCPEGHQSPNCILFIENLTAAQQYGH	578
Db	611	YFCEAOEGSYFREAQHWQLLPEDGIMAE	638
Qy	579	YFCEAOEGSYFREAQHWQLLPEDGIMAE	606
RESULT 2			
ID	088371	PRELIMINARY;	PRT; 393 AA.
AC	088371;		
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)	
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)	
DE	SEMAPHORIN L (FRAGMENT).		
GN	SEMAL		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	MEDLINE; 98389619.		
RX			
RA	LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSEER A.;		
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA		
RT	viruses.";		
RL	GENOMICS 51:340-350(1998).		
RL	DR EMBL; AF030699; G3523117; -.		
FT	NON_TER 393 393		
SO	SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;		
Query Match 52.5%; Score 2226; DB 11; Length 393;			
Best Local Similarity 92.3%; Pred. No. 0.00e+00;			
Matches 301; Conservative 12; Mismatches 12; Indels 1; Gaps 1;			
Db	68	QPEPHTVLFPHEPGSFVWVGGKVIYHFNFPFGKNASVIRVNIWGSTKGCQKQDCGNVI	127
Qy	39	QTEPHTVLFPHEPGSSVWVGGKVIYLFDFPEGNKASVIRVNIWGSTKGCCLDKRCENYI	98
Db	128	TLERGRGILLVCGTNRARKPCSNLNVDSVYMSLGEMKGYAPFSPDENSILVLPFGDEVYS	187
Qy	99	TLERSEGLLACGTNARHPSCNVLNCTVV-PLGEMRGYAPFSPDENSILVLPFGDEVYS	157
Db	188	TIRKQYNGKIPRFRIRGESELYTSDTMQNQPIKATIVHQDQAYDDKIYFFREDNP	247
Qy	158	TIRKQYNGKIPRFRIRGESELYTSDTMQNQPIKATIVHQDQAYDDKIYFFREDNP	217
Db	248	DKNPEAPLNVSRVAQLRCDCQGGESSLSVKWNTFLKAMLCVSDAATNRNFRLODVFL	307
Qy	218	DKNPEAPLNVSRVAQLRCDCQGGESSLSVKWNTFLKAMLCVSDAATNRNFRLODVFL	277
Db	308	PDPGGQRDTRVYGFSPNPNYSACVYSLGIDIDVRFSTSLKGYHMGLSNRPNGMCLPK	367
Qy	278	PDPGGQRDTRVYGFSPNPNYSACVYSLGIDIDVRFSTSLKGYHMGLSNRPNGMCLPK	337
Db	368	KQPIPTTTFQVADRHPEVAQRVEPMG	393
Qy	338	KQPIPTTTFQVADRHPEVAQRVEPMG	363

RESULT		3	
ID	Q64906	PRELIMINARY;	PRT; 653 AA.
AC	Q64906;		
DT	01-NOV-1996	(TREMBREL. 01, CREATED)	
DD	01-NOV-1996	(TREMBREL. 01, LAST SEQUENCE UPDATE)	
DE	01-NOV-1998	(TREMBREL. 08, LAST ANNOTATION UPDATE)	
DE	SIMILAR TO GENBANK ACCESSION NUMBER L26081.		
OS	ALCELAPHINE HERPESVIRUS 1.		
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;		
OC	GAMMAHERPESVIRINAE.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C500;		
RX	MEDLINE; 97201573.		
RA	ENSSER A., FLECKENSTEIN B.;		
RA	ENSSER A., PFLANZ R., FLECKENSTEIN B.;		
RL	"Primary structure of the alcelaphine herpesvirus 1 genome.";		
RT	J. GEN. VIROL. 76:1063-1067(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C500;		
RX	MEDLINE; 97404659.		
RA	ENSSER A., PFLANZ R., FLECKENSTEIN B.;		
RL	SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.		
DR	EMBL; U18243: G1000717: - - - - -		
DR	EMBL; AF005370; G2337970: - - - - -		
DQ	SEQUENCE 653 AA; 73645 MW; CCB53C9 CRC32;		
Query Match	43.1%; Score 1825; DB 14; Length 653;		
Best Local Similarity	47.5%; Pred. No. 0.00e+00;		
Matches	262; Conservative 95; Mismatches 184; Indels 11; Gaps 11;		
Db	95 EPHTVLPHSLNSDDYVGGNTTYLFDFFAHSSNASTALINITSTHTRLSSTCENFRTL 154 : : : : :		
Qy	41 EPHTVLPHEPGSSVMVGRGKYLEDFEPGKNASVRTNIGSTKGSCLDKDCENYITL 100 : : : : : :		
Db	155 LHNOTDGLLAGCTNSOKPCSW-LINNLTTFQLPKGLAPFPSPSSNLVLFDONDITYTI 213 : : : : : :		
Qy	101 LERSSEGLLAGCTNARPSCWNLVGTVP-LGEMRGYAFPSDENSLVLFEGDEVYSTI 159 : : : : : :		
Db	214 NLKXSLGS-HKFRIAGOVELYTSDTAMHRPOFVOATAVHKNESYDDDKIYFFQENSHS 272 : : : : : :		
Qy	160 RK-QEYNGKIPRRIRGESELYTSDTMQNPPQFIKATIVHQDAQYDDKIIYFFREDNDP 218 : : : : : :		
Db	273 DFQGFPTVRPVGVCCSDOGGESSLSVYKWTFLLKARLACVDYDTGRYNELODIFIWQ 332 : : : : : :		
Qy	219 KNPEAPLNVSVAOLCRGDGGESSLSVSKWNTFLKAMLVCSDAATNKFNRLQDVFLP 278 : : : : : :		
Db	333 APENSWEETILGLFLSPWFNVACVFVTVDIDHVEFKTKLKNYHHKLTPRPGQCMMKH 392 : : : : : :		
Qy	279 DPSQQRDTRIVGVSNPWNIYSAVCVISLGIDIKVFTSSLKYHSLSNPREGKCLPQ 338 : : : : : :		
Db	393 QHPVTETFQADRYPEVADPVYQKNNAMEPIIOSKIYYTKLLVYRYEYG-GVFMATIFYL 451 : : : : : :		
Qy	339 QPIPTETFQADRHPVEAQVPEMPGLKTLPLFSKHYQKVAVHRMOASHGETFH-VLYL 397 : : : : : :		
Db	452 TTIKGTHIYVRDSSNTALTALILEINPPQKPAPTONTLLDNTNLKLYNSEWSEVSEP 511 : : : : : :		
Qy	398 TTDRTGTHKKVPEQEQSFAFIMEIQPFPRAAAQIOTMSLDAERKRKLYSSQWSEVQVP 457 : : : : : :		
Db	512 LDLCSVYGNDCFCSMRSDPLCTWYNNTC-S-EK-QRVSVETGCPANRTLSEMCGDHYP 568 : : : : : :		
Qy	458 LDLCEVYGGCHGCLMSRDYFCGWDOGRCLSIYSRSSRVLOSINPAEPH-KE-CPNPKPD 515 : : : : : :		
Db	569 TVVKHQVSIPLLNSYLSCPASNHADYFWTKDGFEKCHKVTHKHNCDILLIANSTATT 628 : : : : : :		
Qy	516 KAPLOKYSIAPNSRYLSCPMESRHATYSWRHKENVYEQSCPGCHQSPNCILFTENLTAQQ 575 : : : : : :		

```
Db 629 NGTHVCMKEDS 640
QY 576 YGHYCEAOEGS 587
RESULT 4
ID Q62177 PRELIMINARY; PRT; 748 AA.
AC Q62177;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN A PRECURSOR (SEM A).
GN SEMA OR SEMA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC -!- EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL
BIRTH.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85990; G854324; -.
DR MGD; MGI:107561; SEMA.
DR PFAM; PF00047; ig; 1.
KW SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
DEVELOPMENTAL PROTEIN.
FT SIGNAL 1 26
FT CHAIN 27 748
FT DOMAIN 586 649
FT IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 748 AA; 82894 MW; A753A8D CRC32;
Query Match 17.4%; Score 737; DB 11; Length 748;
Best Local Similarity 33.7%; Pred. No. 9.61e-154;
Matches 168; Conservative 112; Mismatches 166; Indels 52; Gaps 36;
Db 112 ECMNFVRLHAYNHTLLACRTGAPHTCALNRWATAGTGHASTGPEKLEDGKTPYDP 171
QY 93 DCENXITLLERSEG-LLACGTNARHPSG- ---WNLVNGT-V-P--IGEMRGYAPFSP 142
Db 172 RHRPSPVLGEELYSGV-TADLMGRDFTIFRSLGONPSLRTPEHDSRWLNPEKVKVFWI 230
QY 143 DENSILVLFEGDEVYSTIRKQYNGK-IPFRIRIGESELYT-S-DTV-MONPOFIKAT-I 197
Db 231 PSENPDDDKIYFFRESAVEAAPAMGRMSVSRVGOICRNDLGGORSL-VNKWTTFLKAR 289
QY 198 VHODQAYDDKIYFFREDNPKNPE-APLNVSRVAQLCRDGGESSLSVSKWNTFLKAM 256
Db 290 LVCSVPGVEGDTHFDOLOQVFLLS-SR-DROPTLYAVFSTSSGVFGQSAVCVYSNDVVR 347
QY 257 LVCS--DAATNKNFNRLQDVFLLPDPGSGWRDTRVYGVFSNPWN-Y--SAVCVYSLG 311
Db 348 RAFLGPLPHKEGPTHQWVSQYGRVYPRGCMGPKSTFGTSKDFPDVDVIOFGRNHPLM 407
QY 312 KYFRTS-SLK-G-----YHSSLNPRGKCLPDQ-QPIP-TETF--QVAD--R-HPEV 355
Db 408 YNPVLPMG-S-R-PLFLQVGAGYFTTQIAADRAAADGH-YDVLFGTDVGTVLKVISVPK 464
QY 356 AQRVPEMGLKPTLF-H--SKYHOKVAVHRMQASHGETTFHYLYLTDRGTTHKVVY-P- 410
Db 465 GRPNSEGLLELQVDFEDSAITSQISSKQOQLYVASRAVAQIALHRCALTACRAE 524
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```
QY 411 GEQHSFAFNIMEIQFRRRAAIQTMSLDAERRKLYVSSQWEVSQVPLDLCVYGGGCHG 470
Db 525 CCLARDPYCAMDGSACTREFQAKRRFQIDIRNGDPSTLCSDGSDSHSVLLKVKVLGVES 584
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 471 CLMSRDPCYCGWDGRCISYSSERSVL--QSNPAEPHKECPNPKPDKAPLQKVS LAPNS 528
Db 585 GSAFLECEPRSLQAHVQW 602
QY 529 -RYLSCPMESRHATYSW 545
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
RESULT 5
ID Q62181 PRELIMINARY; PRT; 751 AA.
AC Q62181;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E PRECURSOR (SEM E).
GN SEMAE OR SEME.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC -!- MAXIMUM EXPRESSION BETWEEN DAYS 10-12 WITH MODERATE LEVELS FROM
DAY 13 UNTIL BIRTH.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85994; G854332; -.
DR MGD; MGI:107557; SEMAE.
DR PFAM; PF00047; ig; 1.
KW SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
DEVELOPMENTAL PROTEIN.
FT SIGNAL 1 20
FT CHAIN 21 751
FT DOMAIN 587 649
FT IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 751 AA; 85259 MW; B28D6CFE CRC32;
Query Match 17.1%; Score 725; DB 11; Length 751;
Best Local Similarity 33.0%; Pred. No. 1.35e-150;
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;
Db 167 SENPNVNVSVMINELFSGMYI-DPMGTDAAFRSLTKRMOLRTDOHNSKWLSEPMFVD 225
QY 139 PESPENSLVLFEGDEVYSTIRKQYNGK-IPFRIRIGESELYTS--DTV-MONPOFIK 194
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 226 AHVIDPGTDPNDAKYVFFFKERLTNNRSTKQIHSMIARICPNDTGGORSL-VNKWTTFL 284
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 195 ATIVHQD-QAYDDKIYFFREDNPKNPEAPLNVSRVAQLCRDGGESSLSVSKWNTFL 253
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 285 KARLVCSVTDEGPTHFDELEDVFL-ETDNP-RTTLVYIGITFTSSSVFKGSACVYHL 342
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 254 KAMLVCS--DA-ATNKNFNRLQDVFLLPDPGQWRDTRVYGVFSNPWN-Y--SACVYSL 307
Db 343 SDIQTVNGPFAHKGPNHQLISYGRIPYPRGTCPGCAFTPNMRTTKDFPDVVVTFIR 402
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 308 GDIDKVFRTS-SLK-G-----YHSSLNPRGKCLPDQ-QP-I-PTETF--QVADRP 353
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 403 NHPLMYNISPIHRRPLIVRIGTDYKTKIADVRYNAADG-RYHVLFLGTDRTGTVKVV 461
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 354 EVAQVPEMGLP-KTPLF-H--SKYHOKVAVHRMQASHGETTFHYLYLTDRGTTHKVVY 409
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 462 LPTNSASGELILEEVEFKNHVPITTTMEISSKKOOLYVSSNEGVSQVSLHRCYGTAC 521
```

```
Qy 410 -PGEHSAFAFIMEIQPFRAAAIQTMSLDAERKLYSSQWEVSQVPLDCEVGGC 468
Db 522 ADCCLARDPYCAWDGHSRYPCTKRSRRQDVRHGHNPLTQCRGNLKAVERNAAEIVQY 581
Qy 469 HGCLMSRDPYCGWDGRCISYSSE--RSVLOSINPAEPHKECP--NPKPKAPLQKVSL 524
Db 582 GVRNNTSTFECAPKSPQASIKWLLQKDK 609
Qy 525 A-PSRYYLSCPMESRHATYSWR-HKEN 550

RESULT 6 PRELIMINARY; PRT: 749 AA.
ID Q13214
AC Q13214;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96210603.
RA SERIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT cancer deletion region and demonstrate distinct expression
RT patterns.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR EMBL; U28369; G974284; -.
DR PFAM; PF00047; Ig 1.
SQ SEQUENCE 749 AA; 83121 MW; 6138C5D5 CRC32;

Query Match 16.9%; Score 716; DB 4; Length 749;
Best Local Similarity 33.0%; Pred. No. 3.09e-148;
Matches 167; Conservative 115; Mismatches 170; Indels 54; Gaps 41;

Db 112 ECMNFVKLLHAYNRTHLACGTGAFHPTCAFEVGHRAEPEVLRDLDPGRIDGKGKSPYD 171
Qy 93 DCENITILLER-RSEGLLACGTNARHPSC-WNLV-N-G--TVVPL--G--EM-RGYAPFS 141
Db 172 PRHRAASVLVGEELYSGVA-ADLMGRDFTIFRSLGQRPRLTEPHDSRWLNPEKRFVKVFW 230
Qy 142 PDENSLVLFEGDEVYSTIRKQYNGK-IPFRIRIGESELYT-S-DIV-MQNPOFIKAT- 196
Db 231 IPESNPDDDKIYFFRETAVEAAPALGRLSVSRVGQICRNDVGGQRL-VNKTWTFELKA 289
Qy 197 IVHODQAYDDKIYFFREDNPDKNPE-APLNVSRVAQLCRGDOGGESSLSVSKWNTFLKA 255
Db 290 RLVCSPVGEVDTHFDLODVELLS-SRDH-RTPLLYAVFSTSSIFOGSACVYYSKNDV 347
Qy 256 MLVCS--DAATNKNFNRLQDVFLLPDSPGQWRDTRVYGVFSNPWN-Y--SAVCYISLGD 310
Db 348 RRAFLGPAHKEGPMHOMVSYQGRVPYPRPGMCPKSTFGTFSSTKDFPDDVIOFARNHPL 407
Qy 311 DK--V--F--RTSSL-K--GYHSSLNPRPGKCLPDQ-QPIP-TETF--QVAD--R-HPE 354
Db 408 MYSNVLPTGG-R-PLFLQVGANYTFTQIAADRVAAADGH-YDVLFIGTDVGTVLKVISVP 464
Qy 355 VAORVEPMGPKLTPF-H--SKYHYQKVAHVHMQASHGETHFVLYLTDRGTIHKVVE-P 410
Db 465 KGSRPSAEGLLLELHVFEDSAVTSQISKRHOLYVASSVAQIAHLHCAAHGRVCT 524
Qy 411 -GEOHSAFAFIMEIQPFRAAAIQTMSLDAERKLYSSQWEVSQVPLDCEVGGC 469
Db 525 ECCLARDPYCAWDGACVACRFQPSAKRRRRQDVRNGDPSTLCSGDSRRPALHKKVFGVE 584
Qy 470 GCLMSRDPYCGWDGRCISYSERSVL--QSNINPAEPHKECPNPKPKAPL-QKVS LAP 526
Db 585 GSSAFLECEPRSLQARVETWTFQAGV 610
```

```
Qy 527 NSRYLSCPMESRHATYSWR-HKENV 551

RESULT 7 PRELIMINARY; PRT: 750 AA.
ID Q93018;
AC Q93018;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA DANTE M., WAMBLEY P.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DE [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U73167; G2880035; -.
DR PFAM; PF00047; Ig 1.
SQ SEQUENCE 750 AA; 83034 MW; CBC87E34 CRC32;

Query Match 16.9%; Score 716; DB 4; Length 750;
Best Local Similarity 33.0%; Pred. No. 3.09e-148;
Matches 167; Conservative 115; Mismatches 170; Indels 54; Gaps 41;

Db 113 ECMNFVKLLHAYNRTHLACGTGAFHPTCAFEVGHRAEPEVLRDLDPGRIDGKGKSPYD 172
Qy 93 DCENITILLER-RSEGLLACGTNARHPSC-WNLV-N-G--TVVPL--G--EM-RGYAPFS 141
Db 173 PRHRAASVLVGEELYSGVA-ADLMGRDFTIFRSLGQRPRLTEPHDSRWLNPEKRFVKVFW 231
Qy 142 PDENSLVLFEGDEVYSTIRKQYNGK-IPFRIRIGESELYT-S-DIV-MQNPOFIKAT- 196
Db 232 IPESNPDDDKIYFFRETAVEAAPALGRLSVSRVGQICRNDVGGQRL-VNKTWTFELKA 290
Qy 197 IVHODQAYDDKIYFFREDNPDKNPE-APLNVSRVAQLCRGDOGGESSLSVSKWNTFLKA 255
Db 291 RLVCSPVGEVDTHFDLODVELLS-SRDH-RTPLLYAVFSTSSIFOGSACVYYSKNDV 348
Qy 256 MLVCS--DAATNKNFNRLQDVFLLPDSPGQWRDTRVYGVFSNPWN-Y--SAVCYISLGD 310
Db 349 RRAFLGPAHKEGPMHOMVSYQGRVPYPRPGMCPKSTFGTFSSTKDFPDDVIOFARNHPL 408
Qy 311 DK--V--F--RTSSL-K--GYHSSLNPRPGKCLPDQ-QPIP-TETF--QVAD--R-HPE 354
Db 409 MYSNVLPTGG-R-PLFLQVGANYTFTQIAADRVAAADGH-YDVLFIGTDVGTVLKVISVP 465
Qy 355 VAORVEPMGPKLTPF-H--SKYHYQKVAHVHMQASHGETHFVLYLTDRGTIHKVVE-P 410
Db 466 KGSRPSAEGLLLELHVFEDSAVTSQISKRHOLYVASSVAQIAHLHCAAHGRVCT 525
Qy 411 -GEOHSAFAFIMEIQPFRAAAIQTMSLDAERKLYSSQWEVSQVPLDCEVGGC 469
Db 526 ECCLARDPYCAWDGACVACRFQPSAKRRRRQDVRNGDPSTLCSGDSRRPALHKKVFGVE 585
Qy 470 GCLMSRDPYCGWDGRCISYSERSVL--QSNINPAEPHKECPNPKPKAPL-QKVS LAP 526
Db 586 GSSAFLECEPRSLQARVETWTFQAGV 611
Qy 527 NSRYLSCPMESRHATYSWR-HKENV 551

RESULT 8 PRELIMINARY; PRT: 751 AA.
ID Q99985
AC Q99985;
```


Db 346 DIRWVNGFAHKEGPNQWMPFSGKMPYPRPGTCGGTTPSMKSTKYDPEVINFMRS 405
 || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
 Qy 309 DDKDVFTS-SLK-G--YH-----SS-LPNRPRGKCLPDQ-QP-I-PTETF--QVAD--R- 351
 Db 406 HPLMWCAVPLQRRLPVVRT-CAPYRLTTIAVDQVDSADG-RYEVLFLGTDRGTQVKVI 463
 || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
 Qy 352 HPEVAQRVEPMG--PLKTPLFHSHKYHKQAVAHRMQASHGETHFVLYLTDRTGIHKVV 409
 Db 464 LPKDDDEEELMLEEVFEFKDPAPVKTWTISSKRQQLYVASAGVTHLSLRHCQAAGAAC 523
 || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
 Qy 410 -PGEQHSFAFNIMEIQPFRAAAIQTMSLDARERKLYVSOWSEVSOVPDLDCVYGGC 468
 Db 524 ADCCLARDPYCAWDGQACSRYTASSKKRRSRDQVRHGNIPIROCRGFNSNANKNAVESVOY 583
 || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
 Qy 469 HGCLMSRDPCGWDOGRGIS-IYSSEK-SVLQSINPAEPHKECP--NPXPKAPLQKVSU 524
 Db 584 GVAGSAAFLECPSPQATVKW 605
 : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 Qy 525 A-PNSRYLYSCPMESRHATYSW 545

```

RESULT      11
ID   O88633          PRELIMINARY;             PRT;    754 AA.
AC   O88633;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DI   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SEMAPHORIN IV ISOFORM A.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
NC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   ECKHARDT F., MEYERHANS A.;
RA   "Molecular cloning and expression pattern of a murine semaphorin
RT   homologous to H-sema IV."
RI   SUBMITTED (JUL-1998) TO EMBL/GENBANK/JDBJ DATA BANKS.
DR   ENBL; AF080091; G3377768; -.
SQ   SEQUENCE 754 AA; 85133 MW; EC05BDF3 CRC32;

Query Match      16.1%; Score 682; DB 11; Length 754;
Best Local Similarity 31.8%; Pred. No. 2.41e-139;
Matches 141; Conservative 119; Mismatches 140; Indels 43; Gaps 31

Db 171 PYDPKLDTASALINEELVAGVYI-DEMGTDAIFRTLGKOTAMRTDOYNSRWLNDSPTIH 229
Qy 139 PFPSDENSLFEGDEVSTIRKKYGSK-KIPRERRIGSELYTS--DTV-MQNQFIK 194
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 230 AELIPDSAERNDDKLYFFFRSAB-AQONPAVTARIGRICLNDGDGHCL-VNKKWSTFL 287
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy 195 ATIVHOD-QAYDDKIYYFREDNPDKNPEAPLVNYSVAQLRCGDOGESSLSYKWNTEL 253
Db 288 KARLVCSVPGEDGIETHFELOQVFV-QQTODI-RNPVIYAVTSSGSVFRGSACVCWSM 345
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy 254 KAMLVCS---DAATNKNFRLQDVLLPDPSGQWRDRTVYGVSNP-WNY--SACVCVSL 307
Db 346 ADIRMYENGFPFAHKEGPNQWMPFSGKMPYPRPGTCGGTTFPSMKSTKYDPEVINFMR 405
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy 308 GDIDKVFTS-SLK-G-YH-----SS-LPNRPRGKCLPDQ-QP-I-PTETF--QVAD--R 351
Db 406 THPLMQAVYPLQRRLPVVRT-CAPYRLTTVAVDQVDAADG-RYEVLFLGTDRGTQVKVI 463
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy 352 -HPEVAQRVEPMG--PLKTPLFHSHKYHKQAVAHRMQASHGETHFVLYLTDRTGIHKVV 408
Db 464 VLPKDQEEVEELMLEEVFEFKDPAPVKTWTISSKRQQLYVASAGVTHLSLRHCQAAGA 523
Qy 409 E-PGEQHSFAFNIMEIQPFRAAAIQTMSLDARERKLYVSOWSEVSOVPDLDCVYGGC 467
Db 524 CADCCLARDPYCAWDGQACSRYTASSKKRRSRDQVRHGNIPIROCRGFNSNANKNAVESVO 583
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy 468 HGCLMSRDPCGWDOGRGIS-IYSSEK-SVLQSINPAEPHKECP--NPXPKAPLQKVS 523
Db 584 YGVAGSAAFLECPSPQATVKW 606
```

```

Qy      : | | : | | : || |
524 LA-PNSRYILSCPMESRHATYSW 545

RESULT 12
ID Q90607 PRELIMINARY; PRT; 772 AA.
AC Q90607;
DT 01-NOV-1996 (TREMBREL. 01, CREATED)
DT 01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE COLLAPSSIN.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
CN NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 94006554.
RU LUO Y., RAIBLE D., RAPER J.A.;
RT "Collapsin: a protein in brain that induces the collapse and
RL paralysis of neuronal growth cones.";
RE CELL 75:217-227(1993).
DR EMBL; U02528; G410079; -.
DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 772 AA; 88867 MW; BC6CB9E8 CRC32;

Query Match 16.1%; Score 681; DB 13; Length 772;
Best Local Similarity 31.5%; Pred.No. 4.40e-139; Indels 48; Gaps 38
Matches 146; Conservative 119; Mismatches 150;

Db 166 RGKSPYDPKLITASLLVDGELYSCTA-ADFMGRDFAIFRTLGHHPHPIRTEQHDSRWLNLP 224
||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 135 RGYAPESPDSLNLVFESDEVSTIRKQEYNGK-IPREFRIERGESELYTS--DIV-MQNP 190

Db 225 RFISAHILPSDNPEDDKIYFFRENDAIDEGHTKGATHARIGQICKNDFGHRSL-VNKW 283
||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 191 QFIKATIVHQ-DQAYDDKIYYFFREDNPDKNPEAPLVMSVAQLCRGDQGESSLSVSKV 249

Db 284 TTFLKARLICSVPCNGCIDTHEDELQDVFLM-NSKDP-KNPIYGVGVFTTSSNIFKGSVC 341
||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 250 NTFUKALMVCS--DA-ATNNKENLQDVFLLPPSGQWRTRVIGVSFNPNW-Y--SVC 303

Db 342 MYSTDVRRVRFLGPYAHRDGNPYQGVPRPVPRPGTCPTSGFGFSDTKDLPLDEVIT 401
||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 304 VYSLGDIDKVERTS-SLK-G-YH-S-S-LPMPRGKCLPDQ-QPI-PTEIF-QV-A 349

Db 402 FARSHPWAMPVPFINS-RPIMIKTDVDYQFTQIVLDVRVDAEDGQ-YDMVFIGTIDGTVL 459
||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 350 -DR-HPEVAQRVEPMGPLKTPLFSK--YHYQKVAVHRMQASHGETPHVLVLTTRGTIH 405

Db 460 KVISIPETHWELEEVLEEEMTVFREPTVISAANKISTKOOOLIGSATGVSOJPLHRCDV 519
||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 406 KWE-PGEQEHSFAFNIME-IQPFRRAAAQVTMSLDAERKRKLXVSSQEWVSQVPLDCEV 463

Db 520 YGKACAECCCLARDPYCAWDGSSCSRYPPTAKRRTRRODIRNGDPLTHCSDLQHHDNPSGQ 579
||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 464 YGGGCHCGLSRDPYCWDGDCRCISIIYSSE-RSVL-QSINPAEPHKCPN-PPDKAP-- 518

Db 580 TLEEKIIYGVENSSTFLCSPKSQRAIVY-WQFKQNDDHKVE 621
||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 519 -LQ-KVSLA-PNSRYILSC-PMESRHATYSWR-HKENVEQSCE 556

RESULT 13
ID O88632 PRELIMINARY; PRT; 785 AA.
AC O88632;
DT 01-NOV-1998 (TREMBREL. 08, CREATED)
DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV ISOFORM B.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.

```


Qy	195	ATI VHOD-QAYDDKIYYFFREDPNDPNPEAPLNVNSVAQLCRGDGGESSLSYSKWNTFL	253
Db	319	KARLVCSVPGEDGEGTETHFELOQVEV-QOQTQDV-RNPVLYAVFTSSGSVFRSACVCSVM	376
Qy	254	KAMLVCS---DAATKNFNRLQDELLPDPFSQGMQDTRIVYGVSFN- WNY- -SACVCVSL	307
Db	377	ADIRWVNGFPAKHEGPNYQWMPFSGKMPYPRPGTCPCGGFTPTSMKSTKDYDPDEVINEMR	436
Qy	308	GDIDKVERTS-SLK-G--YH-----SS-LPNPREGKCLDPO-QP-I-PTETF--QVAD--R	351
Db	437	SHPLMYQAVPYLQORRPLVWRT-CAPYRLTITAYDQVDAGDG-RYEVLFGLTDRGTQVKYI	494
Qy	352	HPEAVORVPMG--PLKTLPIFUSKHYQKVAVHRMQASHGETFHVLYLITDRGTIHKVV	408
Db	495	VLPKDDOEMEELMEEVEFKDPAPVKMTITISSKROOLYVASAVGYTHLSHRCAQYAA	554
Qy	409	E-PCEQEHSAFNIETQPFRRAAAQTMSLDAERKLYVSSQWESYQVPLDLICEVYGGG	467
Db	555	CADCCCLARDPYCAWDGQACSRYTASKKRRSRRODVRHGNPIROCGRFNSNANKNVSQV	614
Qy	468	CHGCLMSRDPYCGWDGRCIS-IYSSER-SVLQSIINPAEPHKBCP--NFKPDKAPLQKYS	523
Db	615	YGVAGSAAFLECCQPRSPQATVKW	637
Qy	524	LA-PNSRYILSCPMESRHSYISW	545

Search completed: Thu Jul 8 19:24:40 1999
Job time : 59 secs.

MPERCH_PP

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:28:33 1999; MasPar time 26.94 Seconds
Tabular output not generated. 441.266 Million cell updates/sec

Title: >US-09-041-236-2
Description: (48-606) from US09041236.pap (35 of 45)
Perfect Score: 4170
Sequence: 1 HEPGSSSVWVGGRGKVLFD.....SYFRAQHQLLPEDGIMAE 559

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseg35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.295; Variance 153.116; scale 0.237

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	661	15.9	771	13	Human semaphorin III	4.51e-51
2	620	14.9	477	13	Human collapsin.	4.63e-47
3	608	14.6	775	33	Human semaphorin.	6.87e-46
4	606	14.5	861	22	Mouse CD100 antigen.	1.08e-45
5	599	14.4	861	32	Human semaphorin.	5.18e-45
6	544	13.0	862	22	Human CD100 antigen.	1.15e-39
7	502	12.0	776	32	Rat semaphorin W.	1.32e-35
8	463	11.1	441	13	Vaccinia virus semaphorin	7.45e-32
9	445	10.7	730	13	Grasshopper semaphorin	3.94e-30
10	428	10.3	587	32	Human semaphorin W.	1.65e-28
11	390	9.4	974	33	Human secreted protei	6.71e-25
12	374	9.0	712	13	Tribolium semaphorin	2.17e-23
13	367	8.8	650	13	Drosophila semaphorin	9.91e-23
14	362	8.7	724	13	Drosophila semaphorin	2.92e-22
15	352	8.4	930	32	Human semaphorin Y.	2.54e-21
16	325	7.8	888	25	Human semaphorin Z.	8.39e-19

17	316	7.6	887	25	W19856	Rat semaphorin Z.	5.74e-18
18	296	7.1	929	32	W57259	Rat semaphorin Y.	4.03e-16
19	157	3.8	122	13	R71385	Varola major virus s	2.39e-04
20	99	2.4	591	4	R23006	Protein transcribed f	2.39e-01
21	98	2.4	832	33	W61092	Taq DNA polymerase I	2.82e-01
22	101	2.4	2329	25	W25038	Partial BRCA2 cancer	1.72e-01
23	96	2.3	832	33	W61090	Taq DNA polymerase I	3.91e-01
24	91	2.2	478	1	R04881	Recombinant elastase.	8.71e-01
25	91	2.2	598	9	R48631	Sequence of nuclear r	8.71e-01
26	90	2.2	638	5	R03924	E. coli HSP (dnaK).	1.02e-02
27	92	2.2	832	33	W61091	Taq DNA polymerase I	7.43e-01
28	92	2.2	854	30	W56309	Clas II S-receptor ki	7.43e-01
29	92	2.2	855	6	R29815	S receptor kinase pro	7.43e-01
30	90	2.2	1242	30	W52287	Rattus norvegicus cdo	1.02e-02
31	89	2.1	143	25	W25052	BRCA2 cancer suscep	1.19e-02
32	89	2.1	534	25	W25031	Partial BRCA2 cancer	1.19e-02
33	89	2.1	680	4	R23143	Mutant thermostable D	1.19e-02
34	89	2.1	757	4	R23141	Mutant thermostable D	1.19e-02
35	89	2.1	787	4	R23142	Mutant thermostable D	1.19e-02
36	89	2.1	832	13	R76691	DNA-polymerase REM-T3	1.19e-02
37	89	2.1	832	12	R64272	T. aquaticus DNA-poly	1.19e-02
38	89	2.1	832	1	P90556	Purified native therm	1.19e-02
39	89	2.1	832	33	W61089	Taq DNA polymerase I	1.19e-02
40	89	2.1	832	13	R76690	Taq DNA polymerase RE	1.19e-02
41	89	2.1	833	27	W24211	Cleavase DN nuclease.	1.19e-02
42	89	2.1	833	27	W24212	Cleavase DA nuclease.	1.19e-02
43	89	2.1	833	27	W24213	Cleavase DV nuclease.	1.19e-02
44	89	2.1	3418	26	W19211	Human breast cancer s	1.19e-02
45	89	2.1	3418	26	W23287	Human breast and ovar	1.19e-02

ALIGNMENTS

RESULT 1

ID R71380 standard; Protein; 771 AA.

AC R71380;
DT 21-NOV-1995 (first entry)
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW varola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN WO9507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 60-63; 101pp; English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87446) or varola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA;

Query Match 15.9%; Score 661; DB 13; Length 771;

Best Local Similarity 30.3%; Pred. No. 4.51e-51;

Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;

Db 166 rgksypdpklltasllldgelysgta-adfmgdrfafrtlghhprteqhsrwndp 224

Qy	263	ATNKFNRLQDVFLLPDPSGWRDTRVYGVFSPWNV---SACVVS LGDIDKVF-R----	315
Db	343	mqstaveqstkwrvyngpvtprgaacidsaearanytsslnpdktlqfvdhplmdd	402
Qy	316	--TSSL-----K-G-YHSSLFNP RP GKCL-PDQOPIP-TETFOVADRHPE-V-AQR-VF-	360
Db	403	svtgidnrpxlikkdvnytgivdrqtalqdtfyvmfistdrgalnhkaviltkevvh--v	460
Qy	361	PMGSLKT-P-LFHSKHYQKVAVHVMQASHGETFHVLYLTTRDGTIHKVYVEGEQHSFA	418
Db	461	--laetglrfdfevpllllsskkgkrfyagsgnvvaaplafeckhgs-cedcvlard	517
Qy	419	FNIMEIQPFRRAAIIQMSLDAGER-RKL-LYVSSQWEVSVQVPLDCEVYGGGCHGCLMSRD	476
Db	518	pycawspaikacvtllhqeeas	538
Qy	477	PYCGWDG-R-CISIVSSERS	495

RESULT	6
ID	WI7657 standard; Protein; 862 AA.
AC	WI7657;
DT	24-JUL-1997 (first entry)
DE	Human CD100 antigen.
KW	CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
OS	vaccine.
KS	Homo sapiens.
FH	Key
FT	peptide
FT	/label= Sig_peptide
FT	42..862
FT	/label= Mat_protein
FT	42..553
FT	/label= Semaphorin_domain
FT	554..630
FT	/label= Ig-like_domain
FT	631..733
FT	/label= Stalk_domain
FT	735..752
FT	/label= Transmembrane_domain
FT	753..862
FT	/label= Cytoplasmic_domain
FT	808..815
FT	/label= Phosphorylation
FT	/note= "putative tyrosine phosphorylation site"
PN	WO9717368-A1.
PD	15-MAY-1997.
PF	12-NOV-1996; U18645.
PR	09-NOV-1995; US-556422.
PA	(DAND) DANA FARBEN CANCER INST.
PI	Boussieris V, Freeman GJ, Hall KT, Nadler LM, Schultze JI;
PB	WPI; 97-280982/25.
DR	N-PDB; T60665.
DR	Nucleic acid molecule encoding CD100 antigen - which stimulates
PT	leukocyte response, e.g. B cell aggregation, differentiation,
PT	survival and T cell proliferation
PS	Claim 7; Page 70-72; 135pp; English.
CC	Human CD100 antigen (WI7657) is a novel leukocyte semaphorin-like
CC	protein that stimulates a leukocyte response, including B cell
CC	aggregation, B cell differentiation, B cell survival and/or T cell
CC	proliferation. Its amino acid sequence was deduced from a cDNA
CC	clone (T60565) isolated by COS cell expression cloning. CD100
CC	polypeptides and fusion proteins, nucleic acids, and host cells
CC	expressing CD100 can be utilised in diagnostic and therapeutic
CC	methods involving modulation of B and T cell responses, neuron
CC	axonal growth and immune cell-nerve cell interaction.
SQ	Sequence 862 AA;

Db 107 eclnyrvlqlsatelyvcgtagnafqpcadchlnltsfkfkgknedkgrocpdpahsyts 166
 : : ||| : | : |||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Qy 93 DCENYITLIERSE-GLLACGTTNARHPSCWNLVNGTVVLPG--EM-RGYAPSPDEN-SL 147
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Db 167 vmvdg-elysgts-ynfigsepilrnsnshspltevaipwnpepsfvfadvirksdpap 224
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Qy 148 VLFEQDEVISTRKQYNGKIPRFRIRGESELVTSDTV--MQNPQFIKATIVHQ--D-- 201
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Db 225 dgeddrvffftevsveyefvrlipriarvcxkdgggirtiq-kwtsflkarlicsr 283
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Qy 202 QAYDDKIYYFREDNDKNPEAPLNSVRAQLCRGDQGESSLSVKWNFLKAMLVCSO 261
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Db 284 pdsgrlvnlrdvflrsp-gl-kvpvfyalftpqlnvnlglsavcaynlstaeefshok 341
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Qy 262 AATNKFNELQDVLFLLDPDSGOWRDTRYGVGSFPWNY---SVCVYSLGIDIKVF--R 315
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Db 342 ymqstveqshckwryngpvkprrgcacidearaanytsslnldktlqfvkdhplmd 401
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Qy 316 ----TS---S-LK-G-YHSSLPNPARGKCL-PDOQPPI-TETFOVADRHE-V-AQR-VE 360
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Db 402 dsvtidnprlikdvytiqtivdrtaqldgtydmfvstdrgalhkai--s-Lehav 458
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Qy 361 -PMGLPKT-P-LFHSHYHKQVAVHRMOASHGFHVLYLTDRGIHKVVPEGOEHSF 417
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Db 459 hi-ieetqlqfqepvqtllleskknrfyvagsnvgaaplafcgkght-ceedvlar 516
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Qy 418 AFNIMEIQPRFAAAITMSLDIAER-RKL-YVSQWEVSQVPDLCEVYGCGGCHGLMSR 475
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Db 517 dpycavspptatcvahlhtespsrgllqemsqdasvcpd-ks-kgyqrhffkhggtael 574
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Qy 476 DPICGWQD--GRCIISYSSRSVLQSINP-AEPHKCPNKPDKAPLKQVSLAPNSRYLV 532
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Db 575 kscsqkanlarvwkfngykaespkyglmrknllif-nlsgdsqvgvqcisee 628
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :
 Qy 533 SCPMESRHATYSWRKHENVQSCPE--GHQSPNCILFIENLTAAQQYGHVFCFAQE 585
 : : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :| : | : ||| :

RESULT 7

ID	W51313 standard; Protein; 776 AA.
AC	W51313;
DT	08-SEP-1998 (first entry)
DE	Rat semaphorin W
KW	Rat; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KX	Immunosuppressant; gene therapy; diagnosis; research reagent.
OS	Rattus norvegicus.
PN	WO9815628-A1.
PD	16-APR-1998.
PF	03-OCT-1997; J03549.
PR	09-OCT-1996; JP-287636.
PA	(SUMU) SUMITOMO PHARM CO LTD.
PI	Kikuchi K, Kimura T;
WP	WPI: 98-261015/23.
DR	N-PSDB: V07279, V07280.
PT	Nerve extension inhibitor protein semaphorin W - is useful as therapeutic drug and diagnostic and research reagent
PS	Claim 1; Page 60-64; 90pp; Japanese.
CC	The present sequence represents rat semaphorin W. Semaphorin W and its derivatives are nerve extension inhibitors which are useful as antiallergic, immunosuppressant and anticancer agents. The DNA encoding semaphorin W can also be used in gene therapy, e.g. using a viral vector. The proteins, peptides, DNA and antibodies which recognise the protein or peptides, can be used as diagnostic or research reagents. Semaphorin W can be used as a screen for semaphorin W antagonists with possible therapeutic use.
SQ	Sequence 776 AA:

Db 135 aivnashlltctfadbpcgvidvssfqgverlesgrgkcbfepaqraaavmagvlyt 194
QY 102 ERRSEG-LLACGTNARHPSCWNL-VNG--TVVPLGEMRGYAPFSPDENSILVLFEGDEVYS 157
Db 195 atkv-nflgteplisarvgraedwrtetlsswlnapafvaamvlspeawegdedgdeif 253
QY 158 TTRKQYNGKIPFRFRIRGESLYR-SDIV---MNPQFIKATIV-H-Q--QDAYDDKIY 209
Db 254 ffftetsrvldseyikvrvarvcagdlgrgkrtlq-qwtwflkadtllcpgehrasg 312
QY 210 YFFREDNPKNEAPLNVSRVAQLCRGDOGGESSLVSQKNTFLKAMLVCSDAATNKNFN 269
Db 313 vlqamaelrpqga-gtplifyfifsqwegaaisavcafrpdidravlngpfrelkhden 371
QY 270 RLQDFVLLPDPGQWRDRTRVYGFSPNWN---YSACVYSLGDIDKVFRTS--SLK-GYH 323
Db 372 rglpymdnvcprrpgeciannmklgqfsslsldprvltfirdhplmdrpfpadgrpl 431
QY 324 SSLP---N---PRGKCLDQOQIP--TETQVADRHE-V-AQR-VZ-PMGPLKT-PL 369
Db 432 lvttdaylrvvahrvtslsgkeydvlylgtedghlhravrigaq-ls-vledlal--fp 487
QY 370 FHSK-YHYOKVAVHRMQASHGETFHVLYLTDRGTIHKVPEGEQESHFAFNIMEIQPFR 428
Db 488 epqvvesmklhbw--llvgshtextvntscnrglqs-csecillaqdpvcawsfldac 544
QY 429 RAAAIQTSLDAERKRLYSSQWESQVPLDLCEYVGCGCHGLMSRDPYCGWD-Q-QRC 486
Db 545 vahagehgmqvadiesadvsslcpegehpvfvprvatghvlpcespsawascvwh 604
QY 487 IISYSSERSVLOSINPAEPHKECPKPKDKAPLQKVSUAPNRSYLYLSCPMESRHATYSWR 546
Db 605 qpsgv 609
QY 547 HKEV 551

RESULT 8
ID R71381 standard; Protein; 441 AA.
AC R71381;
DT 21-NOV-1995 (first entry)
DE Vaccinia virus semaphorin IV protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW varicella major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Vaccinia virus.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; Q87443.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 55-67; 101pp; English.
CC The sequence of the vaccinia virus semaphorin IV protein.
CC The gene sequence was isolated as the A39R open reading frame sequence
CC from varicella, based on sequence homology searches of a database with the
CC grasshopper, Tribolium and Drosophila semaphorin sequences. The proteins
CC encoded by the grasshopper semaphorin I (Q87443), human semaphorin III
CC (Q87444-5), vaccinia virus semaphorin IV (Q87446) or varicella major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and

CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 441 AA;
Query Match 11.1%; Score 463; DB 13; Length 441;
Best Local Similarity 33.9%; Pred. No. 7.45e-32;
Matches 87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;
Db 115 lvcgttnnnpkckw-ldgsdddpkhrgrgyapqyskvliishngc-vlsdinisk-eg-i 170
QY 109 LACGTNARHPSCWNLVNGTVVPLGEMRGYAPFSPDENSILVLFEGDEVYSTIRKQYNGKI 168
Db 171 krwrfrdpccgydlytadnvpikdg-lrgafvdkdgttd-kvylftdtgskr--i-vk 225
QY 169 PRFRIRGE-S-ELTSDTVWQNPQFIKATIVHQDAVDKRIYFFREDNPKNEAPLN 226
Db 226 lpylaqmcIndeggpsslsrhwstflkvelec-did-grsy-r-q-ihstrtktd-nd 279
QY 227 VSRVAQLCRGDOGGESSLVSQKNTFLKAMLVCSDAATNKNFNRLQDYVLLPDPGQWRD 286
Db 280 tilyvffdspsyskalcstysmntlqskfstsklegkytkqlpsagsgiclpagkvvphttf 339
QY 287 TRVYGFSPNWNYSACVYSLGDIDKVFRTSLSKGYHSSLPNRPFGKCLPDDQPIPTET 346
Db 340 eviekynvlddiikpls 356
QY 347 QVADRHEVAQRVEPMG 363

RESULT 9
ID R71379 standard; Protein; 730 AA.
AC R71379;
DT 21-NOV-1995 (first entry)
DE Grasshopper semaphorin I protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW varicella major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Grasshopper sp.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; Q87441.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 1; Page 68-72; 101pp; English.
CC The sequence of the grasshopper semaphorin I protein. The proteins
CC encoded by the grasshopper semaphorin I, human semaphorin III (Q87442),
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87446) or varicella major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 730 AA;
Query Match 10.7%; Score 445; DB 13; Length 730;
Best Local Similarity 28.4%; Pred. No. 3.94e-30;
Matches 124; Conservative 114; Mismatches 153; Indels 46; Gaps 38;
Db 106 dcqnyrvrlakidddrictgnaykplcrhyalkdgyvvekegyegrlcfdpdhst 165
QY 93 DCENITILLERSEG-LLACGTNARHPSCWN--LVNGI-VVPLG-EMRGYAPFSPDENSIL 147
Db 166 alysegqlysatv-adfsgtdp-li-yrg-p-lrtersdl-k-qinapnfvn-tmeyndf 217

```
QY 148 VLFEGDEYVSTIRKQYNGKIPRFRIRGESELYTSDTVMQNPQFIKATIVHODQAYDDK 207
Db 218 lffretaveyincgaiky-srvavckhdkg-phaggdrwtsfiksrlncsvpadyp 275
QY 208 IYFFREDNP-D-KNEAPLVNRSVAQLCRGQGESSLSVSKWNTFLKAMLVCSDAATNK 266
Db 276 fyfneigatsdiegnvggq-vekliygvtftvnsigsgsvacafsmksilesfdgpkfe 334
QY 267 -NFNRLQDVF-LLPDP-SGQWRDIRVYGVFSNPWN-Y--SACVYSL-G--D-IDKVFRT 316
Db 335 qetmnsnwlavpslkvppeprgqcvnderltpdvsnfvkshtlmdvpaafft-rpili 393
QY 317 -SSLKG-YHS--SL--PNRPCKLCPDQPIPTETTFQVADRHPVEAORVPMGPKLTPLF 370
Db 394 rislqyrfkiklavdgqvrtpdgkaydvlfictddgkvikalnsasfssdtdvsvleel 453
QY 371 H-S-KYHQRKAV-HRMQASHGETFHVLYLTDRGTTHKVVPEG--BOEHFAFNIME-I 424
Db 454 qvlppgvknlyvvrmdgddsklvvssddelailkhrsgskitncrcvslqdpypca 513
QY 425 QPFRAAAIOTM---SLDAERRKLYVSSQWEVSQVPLDLCVEYG-GGCHGCLMSRDPYCG 480
Db 514 wdnvelkctavgsdpws 530
QY 481 WQD-G-RCISYISERS 495

RESULT 10
ID W51314 standard; Protein: 587 AA.
AC W51314;
DT 08-SEP-1998 (first entry)
DE Human semaphorin W.
KW Human; semaphorin W.
KW Immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PN W09815628-AI.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMU ) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI; 98-261015/23.
DR N-PSDB; V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC anti-allergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 10.3%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 1,65e-28;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgdelyffttstafdsyerikyprvarvcagdlgrktliq-grwtftlkadllcp 115
QY 201 DOAYDDKIYFFREDNPDKNPEAPLVNRSVAQLCRGQGESSLSVSKWNTFLKAMLVCS 260
Db 116 gpehgrassvlqdvavrlpelga-gtptfygifsqwegatisavcafrpddirtvngp 174
QY 261 DAAATKNFNRLQVFLLPDPGQWRDRVYGVFSNPWN---YSACVYSLGIDKVFRTS 317
Db 175 frelkdcnrglpvvdvpprgqecitnnmkrlrhfgsslsldrvltfirdhplndrp 234
QY 318 --SLK-GYHSSLP--N---PRFGKCLPDQOPIP--TETFOVADRHP-E-V-AOR-VE-P 361
Db 235 vfpadghplvttdtaylrvrvahrvtlsqgkdvlylgtedghlhravrigaq-ls-vl 292
```

```
QY 362 MGPKLT-PLFHSK-YHQKVAVHRMQASHGETFHVLYLTDRGTTHKVVPEGGEHSFAE 419
Db 293 edlal--fpqpqvemkml-yhsw-llvgstevtqvnttncgrlqs-csecillaqdpvc 347
QY 420 NMEIQPFRAAAIQTMTSLDAERRKLYVSSQWEVSQVPLDLCVEYGGCHGCLMSRDPYC 479
Db 348 awsrldccvahagehrglvgqdesadvslcpkeperpvtovpvatauhvvlpcsp 407
QY 480 GWD-Q-GRCSYISERSVQLQSINPAEPHKECPNPKDPKAPLOKVLAPNSRYLSCPME 537
Db 408 sawascvwhqpsgv 421
QY 538 SRHATYSMRHKENV 551

RESULT 11
ID W64221 standard; Protein: 974 AA.
AC W64221;
DT 06-OCT-1998 (first entry)
DE Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN W09827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GEMY ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavalie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 98-362774/31.
DR N-PSDB; V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17j; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 9.4%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 6.71e-25;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechmfikvllknddaalfvcgtfnafnpscrnykmdtlepfdefsgmarcpdydahan 172
QY 91 KRDCENYIT-LLERRSEGLACGTNARHPSCWNLVNTVVPLG-EMRGYA--PFSDENS 146
Db 173 valfadgklysatvt-dflaidaviysl-gesp--clrtvkhdkskwkepyfvqadv 228
QY 147 LVLFEGDEVYSTRKQYNG-KIPFRFRIRGESELYTSDTVMQNPQFIKATIVHODQAYD 205
Db 229 dyiyffireaveyntmgkvfvbrvacvckndmgsgvrllekqwtfsfkarlncsvpgds 288
QY 206 DKIVYFPREDNPDKNPEAPLVNRSVAQLCRGQGESSLSVSKWNTFLKAMLVCSDAATN 265
Db 289 hfyfnlqavtdviring--rdv-vlatfstpynspgsvacaydmldiasvftgrfkeq 345
QY 266 K-NFNRLQVFLLPDPGQWRDRVYGVFSNPWN-Y--SACVYSLGIDKVF--RTSSL 319
```


Db 346 kspdstwtpvderypkprpgccagsssleryatsnefoddtnlnfikthlmdeavpsif 405
Qy 320 KQYHSS---LNPREGKCLPD--QQPIPTETTFQVADRPEVA-Q--RVPE-N--G-P-L- 365
Db 406 nrpwlrmvryrltkiavdtaagy-qhntvvlfgsekglilkflaigsgfndslf 464
Qy 366 KTLPLF-HS--KYHYQKVAHRMQASHGETFHVLYLTDRGTIHKVVEP-GEDE--HSAF 419
Db 465 leemsvynseksgydvdkrimgmqldrasslyvafstcvikpgrcerhgk-ckkt 523
Qy 420 -NIMEI-OPFRA-AAIQ---TMSLDAER--RKLYVSSQWESQVPLDLCELYGGGCH-G 470
Db 524 ciasrdpvcgwikegacshlspnrlfcdiergn 560
Qy 471 CLMSRDPYCGW--DOGRCSISYSSERSVL-OSINPAE 504

RESULT 12
ID R71384 standard; Protein; 712 AA.
AC R71384; 1995 (first entry)
DE Tribolium semaphorin I protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Tribolium sp.
PN WO9507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; 087446.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 85-88; 101pp; English.
CC The sequence of the beetle Tribolium genomic DNA. The gene was
CC isolated by PCR using Tribolium semaphorin III (087442), the
CC grasshopper semaphorin I (087441), human semaphorin III (087443),
CC vaccinia virus semaphorin IV (087443), Drosophila semaphorin I and II
CC (087444-5), Tribolium semaphorin I or variola major (smallpox) virus
CC semaphorin IV (087447) genes were used to generate a series of peptides
CC (R70370-R70418), which retain semaphorin receptor binding activity. The
CC semaphorin derived or semaphorin receptor derived peptides are potent
CC modulators of nerve cell growth, immune responsiveness and viral
CC pathogenesis. They can be used in diagnosis and treatment of neurological
CC disease and neuro-regeneration, immune modulation and diagnosis and
CC treatment of viral and oncological infection and diseases.
SQ Sequence 712 AA;

Query Match 9.0%; Score 374; DB 13; Length 712;
Best Local Similarity 27.4%; Pred. No. 2.17e-23;
Matches 139; Conservative 123; Mismatches 191; Indels 55; Gaps 42;

Db 57 tsilvggrnrynlisfclserkggrldwps-sdshgqilcklkgktdcdqnyirilyss 115
Qy 53 SSVWVGGRKVV-L--PDFPEKNASRTVNIQSGKGS-C-L-DRD--CENYITLLERR 104

Db 116 epgkivlctnsykpclrtayafekgylvekevegiglcypnhehnstsvsyng-qlfsa 174
Qy 105 SEG--LLACGTNARHPSCNVLV--NGT-VVPLG-EMRGVAPFSPDENSEL-VLFEQDEVYST 158

Db 175 tv-afdsqgdpliypartel-sdlklnapnfns--v-----aygdiiffyretave 226
Qy 159 IRKOBYNGKIPFRIRGESELYTSDTYVQNPFQTKATIVHQDQYDDKIYFFREDNPD 218

Db 227 ymncgkvlv-ervarvckdkkg-phqsrdrwtstfklrhlncspigeypfyfdelqstsd 284
Qy 219 -KNPEAPLNVSRVAQLCRGDGGESSLSVSKWNTFLKAMLVCSDAATNK--NFNRLQDVEL 276

Db 285 ivegrynsddakkilygilttvpvnaigsgsaicaycmadilrvfegsfkhetinswlpv 344
Qy 277 LPDPSGQWRDTR--VYGVSNPNW-Y--SAVCVYSLGIDKVF-----R-TSLKG-Y-H- 323
Db 345 pqlnvpeprpgqcvrdsvilpdkvfnfikthlmed-vpalfg-kpvlrvslqyrfai 402
Qy 324 S-SL-PNRPREGKCLPDQQPIPTETTFQVADRPEVAQRPEPMGLKTPLEH-S-KHYQKV 379
Db 403 tvdpqvktnnqyldvlygtddgkvikavnipkrhakallyrkyrtsvhphgapvklk 462
Qy 380 AVH-RMQASHGETFHVLYLTDRGTIHKVVE-PCQEHSAFNIEMEIQPFRAAAIQIMS 437
Db 463 iapyygkvvvvgkdeirlanlnhcas-ktckdcvelgdphcawdanknlcvsidtvsy 521
Qy 438 LDAERRKLYVSSQWESQVPLDLCEVYGGGCHGLMSRDPYCGWD--OGRCSISYS-SE- 493
Db 522 rfliqdvrvrgdnk-cwspqtdkktvik 548
Qy 494 RSVLQSNPAEPHKECPNPKPKAPLQK 521

RESULT 13
ID R71382 standard; Protein; 650 AA.
AC R71382; 1995 (first entry)
DE Drosophila semaphorin I protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Drosophila sp.
PN WO9507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; 087444.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 74-77; 101pp; English.
CC The sequence of the Drosophila semaphorin I protein. The gene was
CC isolated by PCR using primers based on sequence homology between the
CC complete grasshopper and partial Tribolium and Manduca (moth) semaphorin
CC gene sequences. The products gave two different sequences, each of which
CC was used to obtain its respective complete sequence: semaphorin I and II
CC (087445). The proteins encoded by the grasshopper semaphorin I (087441),
CC human semaphorin Iir (087442), vaccinia virus semaphorin IV, Drosophila
CC semaphorin I and II, Tribolium semaphorin I (087446) or variola major
CC (smallpox) virus semaphorin IV (087447) genes were used to generate a
CC series of peptides (R70370-R70418), which retain semaphorin receptor
CC binding activity. The semaphorin derived or semaphorin receptor derived
CC peptides are potent modulators of nerve cell growth, immune
CC responsiveness and viral pathogenesis. They can be used in diagnosis and
CC treatment of neurological disease and neuro-regeneration, immune
CC modulation and diagnosis and treatment of viral and oncological infection
CC and diseases.
SQ Sequence 650 AA;

Query Match 8.8%; Score 367; DB 13; Length 650;
Best Local Similarity 28.4%; Pred. No. 9.91e-23;
Matches 124; Conservative 104; Mismatches 158; Indels 51; Gaps 38;

Db 3 dcqyvirimvpsprlfcvgtstnsfrpmctnyilsdsnytleatkgavcypdhrnst 62
Qy 93 DCENYITLLERRSEG--LLACGTNARHPSCWN-LVNGTVVPLGEMR-GYA--PFSPDENSEL 147

Db 63 svladnelystg-adfsgsdpiiyreplteqeqdsils-lnapnfv-ssft-qgd-f--- 114
Qy 148 VLFEQDEVYSTIRKOBYNGKIPFRIRGESELYTSDTYVQNPFQTKATIVHQDQYDDK 207

QY 424 -IQPF---R----RAA--A--IQTMSLDAERRKLYVSSQWEVSQVPLDLCEVYGGCH-G 470
Db 529 clasqdpvcgwhsrgcvdirgsggtdvdqagngesmhgdcdq 572
QY 471 CLMSRDPYCGWDOGR-CISYSS- RSVLQSI NPAE-PHKECPN 511

Search completed: Thu Jul 8 19:29:12 1999
Job time : 39 secs.

THIS PAGE BLANK (USPTO)

WPSRL

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:27:43 1999; Maspar time 25.79 Seconds
Tabular output not generated. 868.442 Million cell updates/sec

Title: >US-09-041-236-2
Description: (48-606) from US09041236.pap (35 of 45)
Perfect Score: 4170
Sequence: 1 HEPGSSVWVGGRGVYLF.....SYFRAQHQLLPEDGIMAE 559

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1.pir1 2.pir2 3.pir3 4.pir4

Statistics: Mean 48.924; Variance 86.181; scale 0.568

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1771	42.5	653	2	T03102 semaphorin homolog A3	0.00e+00
2	737	17.7	748	2	I48744 semaphorin A - mouse	7.80e-135
3	725	17.4	751	2	I48748 semaphorin E - mouse	4.57e-132
4	716	17.2	749	2	G01856 semaphorin V - human	5.42e-130
5	688	16.5	753	2	G02173 semaphorin III family	1.50e-123
6	681	16.3	772	2	A49069 semaphorin - chicken	6.06e-122
7	667	16.0	666	2	I58169 semaphorin III - mouse	9.84e-119
8	666	16.0	772	2	I48747 semaphorin D - mouse	1.67e-118
9	661	15.9	771	2	D49423 semaphorin III precursor	2.33e-117
10	618	14.8	782	2	I48746 semaphorin C - mouse	1.55e-107
11	522	12.5	834	2	S66498 M-sema F protein prec	7.96e-86
12	480	11.5	760	2	I48745 semaphorin B - mouse	1.92e-76
13	470	11.3	403	2	E42521 A39R protein - vaccin	3.21e-74
14	463	11.1	441	2	S29921 hypothetical protein	1.14e-72
15	447	10.7	730	2	JH0798 fasciclin IV precursor	3.94e-69
16	448	10.7	1074	2	JC5928 semaphorin F precursor	2.37e-69
17	385	9.2	711	2	A49423 semaphorin I precursor	1.42e-55
18	367	8.8	656	2	B49423 semaphorin I - fruit	1.09e-51
19	362	8.7	724	2	C49423 semaphorin II precursor	1.29e-50
20	293	7.0	295	2	JQ1775 Sall9R protein - vacc	4.78e-36
21	196	4.7	1884	2	JC4975 plexin 2 precursor -	9.15e-17
22	170	4.1	1872	2	JC4976 plexin 3 precursor -	5.47e-12
23	171	4.1	1905	2	I31553 Plexin - African claw	3.62e-12

24	168	4.0	1894	2	JC4980	plexin 1 precursor -	1.25e-11
25	157	3.8	122	2	JQ1845	I4R protein - variola	1.07e-09
26	157	3.8	122	2	H36852	A43R protein - variol	1.07e-09
27	157	3.7	142	2	JQ1776	SalFIR protein - vacc	2.39e-09
28	112	2.7	775	2	E70320	polyribonucleotide nu	1.89e-02
29	105	2.5	227	2	S09922	hypothetical protein	1.87e-01
30	105	2.5	235	1	Q0BEC9	HXLF4 protein - human	1.87e-01
31	103	2.5	405	2	B69064	conserved hypotheticala	3.52e-01
32	102	2.4	275	2	A34866	T-cell surface protei	4.82e-01
33	99	2.4	275	2	S08484	T-cell alloantigen RT	1.22e+00
34	99	2.4	460	2	S35772	translation elongatio	1.22e+00
35	99	2.4	563	2	S78224	virulence-associated	1.22e+00
36	99	2.4	591	2	S26565	virulence-associated	1.22e+00
37	99	2.4	591	2	S09498	virulence-associated	1.22e+00
38	99	2.4	591	2	S22664	virulence-associated	1.22e+00
39	99	2.4	593	2	S15215	virulence-associated	1.22e+00
40	99	2.4	597	2	A46050	thyroid/steroid recep	1.22e+00
41	98	2.4	970	2	C57282	ankyrin-related prote	1.65e+00
42	98	2.4	979	2	B57282	ankyrin-related prote	1.65e+00
43	99	2.4	1375	2	JC5148	hepatocyte growth fac	1.22e+00
44	98	2.4	1786	2	A57282	ankyrin-related prote	1.65e+00
45	97	2.3	630	2	T00352	hypothetical protein	2.23e+00

ALIGNMENTS

RESULT	1
ENTRY	T03102
TITLE	semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM	#formal name alcelaphine herpesvirus 1
DATE	24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
ACCESSIONS	24-Mar-1999
REFERENCE	T03102
#authors	Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal	J. Virol. (1997) 71:6517-6525
#title	Primary structure of the alcelaphine herpesvirus 1 genome.
#accession	T03102
#status	preliminary; translated from GB/EMBL/DBJ
#molecule_type	DNA
#residues	1-653 #label ENS
#cross-references	EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY	#length 653 #molecular-weight 73645 #checksum 5501
Query Match	42.5%; Score 1771; DB 2; Length 653;
Best Local Similarity	46.8%; Pred. No. 0.00e+00;
Matches	255; Conservative 95; Mismatches 184; Indels 11; Gaps 11;

Db	102	HSLNSSDYYGGNTIYLFDFAHSSNASTALINTTSTHNRSLSTGCENFTLLHNOTDG	161
Qy	48	HEPGSSVWVGGRGVYLFDFPEGKNASVVRTNIGSTKGCCLDRDCENYITLLRSE	107
Db	162	LLAGTNSQKPSCW-LINLLTQFGPKGLAPESPSSGNLVLFQDNDTSTINLYKSL	220
Qy	108	LLAGTWARHPSCNWLNVGVVWP-LGEMRYAPSPDENSLVLFEGDEVSTIRK-QEYN	165
Db	221	GS-HKFRRIAGQVELYTSMTAMHRPQVQATAVHKNESYDKIYFFQENSHSDFKQFP	279
Qy	166	GKIPFRFRIRGESELYTSDVMQNPQIKATIVHQDAQYDDKIYFFREDNPKNPEAPL	225
Db	280	TVPRVGVCSDDGGSSLSVYKWTTELKARLACVDYDTGRIVNELQDIFIWAPENSW	339
Qy	226	NVSRVAQLCRDGGESSLSVKWNTLKAFLVCSDAATNKNFNLQDLVLLPDPSPQWR	285
Db	340	ETLYGLFLSPWNFSVCVTVTDIDHVKTSKLNKHHKLPTRPCOCKMKNHQHVET	399
Qy	286	DTRVYGVFSMPNVSVCVYSLGDIKVFRTSSLKGVHSSLPNRPGRKCLPQPIPT	345
Db	400	FOVADRYPEVADPYQKNNAMFPIQSKYIYTKLLVYRVEYG-GVFWATIFYLTIKGTI	458
Qy	346	FOVADRHPEVAQRVEPMGPKLTPLFHSHKYQKVAVHRMQASHGETFH-VLYLTDRTGI	404
Db	459	HIYRVYEDSNTTALNTLEINPFQKPAPIONILLDNTNLKLYVNSEWSEVSEVLDLCSVY	518

QY 405 HKYVEFGEQEHSAFNIQEPFRAAAIQTMSLDAERKLYVSSQWEYSQVPLDLCEYV 464
Db 519 GNDCFSCFMSRDLPLCTWYNNTC-S-FK-ORVSVETGPGANRTLSEMGDHYAPTIVVKHOV 575
QY 465 GGCCHCLMSRDPYCGWDGRCISIIYSRSRVLSQSNPAEPH-KE-CPNPKDKAPLQKV 522
Db 576 SIIPLSNSYLSCPAVSNHADYFTWKDGFTEKRCRCHVTHKNDICILLIANSTTATNGPHVCN 635
QY 523 SLAPNSRYLSCPMESRHATYSWRHKNENVEQSCPCGHQSPNCILFIENLTAQQYGHYFCE 582
Db 636 MKEDS 640
QY 583 AQEGS 587

RESULT 2 I48744 #type complete
ENTRY semaphorin A - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
DATE 04-Sep-1998

ACCESSIONS I48744
REFERENCE I48744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48744

GENETICS semA
#gene semA
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 748 #molecular-weight 82894 #checksum 9017

Query Match 17.7%; Score 737; DB 2; Length 748;
Best Local Similarity 33.7%; Pred. No. 7.80e-135;
Matches 168; Conservative 112; Mismatches 166; Indels 52; Gaps 36;

Db 112 ECMNFVRLHAYNHTLACRTGAFHPTCALMRWATAGGTHAGTSGPEKLEDGKGTPTDP 171
QY 93 DCENYITLLRRSEG-LLACGTNARHPC----WNLVNGT-V-V-P--LGEMRGYAPFSP 142
Db 172 RHRPPSVLVEELYSGV-TADLMGRDFTIFRSLGQNPSLRTPHDSRWLNPEKFKVFWI 230
QY 143 DENSLVLFEGDEVYSTIRKQYNGK-IPRFRIRIGESELYT-S-DTV-MQNPOFIKAT-I 197
Db 231 PESENDDDKIYFFRESAVEAPANGRMSVSRVQICRNDLGGQBSL-VNKNWTTFLKAR 289
QY 198 VHODQYDDKIYFFREDNPKNPE-APLNVSRVAQLCRDQGGESSLVSKWNTEFLKAM 256
Db 290 LVCSVPGVEGDTHFDQLQDVLLS-SR-DROTPLLXAVRSTSGVFGSAVCVYSNDVR 347
QY 257 LVCS--DAATNKNFNRLQDVLLPDPGQWRDTRYGVGSFNPNW-Y--SAVCVYSLGDD 311
Db 348 RAFGLPLPKHEGTHQWYSQGRVPYPRPGMPCSKTGTGTSKDFPDVDFQGRNHPLM 407
QY 312 KVFTS-SLK-G-----YHSSLNPRPGKCLPDQ-QPIP-TETF--QVAD--R-HPEV 355
Db 408 YNPVLPNGG-R-PLFLQVAGNYFTTQIADRVAAAGH-YDVLFIGDTGTVLKVSVPK 464
QY 356 AQRVPEMPLKTLF-H--SKYHQVAVHRMQASHGETFHVLYLTDRGTHHKVVE-P- 410
Db 465 GRPNSEGLLELQVFEPSAAITSMOISKROOLVKSRAAQAIALHRCALTGRACAE 524
QY 411 GEGHSAFNIQEPFRAAAIQTMSLDAERKLYVSSQWEYSQVPLDLCEYVGGCHG 470
Db 525 CCLARDPYCAWDGSACTRFOPTAKRFRRODINGDPSTLCSDGSDSHVLLKVKVLGVS 584

QY 471 CLMSRDPYCGWDGRCISIIYSRSVLT--QSINPAEPHKECPNPKDPKAPLOKVS LAPNS 528
Db 585 GSAFLECEPRSLQAHQVW 602
QY 529 RYILSCPMESRHATYSW 545

RESULT 3 I48748 #type complete
ENTRY semaphorin E - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
DATE 04-Sep-1998

ACCESSIONS I48748
REFERENCE I48748
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48748

GENETICS semE
#gene semE
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 751 #molecular-weight 85259 #checksum 8961

Query Match 17.4%; Score 725; DB 2; Length 751;
Best Local Similarity 33.0%; Pred. No. 4.57e-132;
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;

Db 167 SPNPVNVYSVMINEELFSGMYI-DFMGTDAAIFRSLTKRMOLRTDOHNSKWLSEPMFVD 225
QY 139 PFPSPDENSILVFEDEVYSTIRKQYNGK-IPRFRIRIGESELYTS--DTV-MQNPOFIK 194
Db 226 AHVIPDGDTPNDKAVYEFKRLTDNNRSTKQIHSMIARICPNDTGQBSL-VNKNWTTFL 284
QY 195 ATIVHQD-QAYDDKIYFFREDNPKNPEAPLNVSRVAQLCRDQGGESSLVSKWNTEFL 253
Db 285 KARLYCSVTDEGPTHEDLEDFELL-ETDNP-RTTLVYIGFTTSSSVFKGSACVYVHL 342
QY 254 KAMLVCS--DA-ATNKNFNRLQDVLLPDPGQWRDTRYGVGSFNPNW-Y--SAVCVYSL 307
Db 343 SDIQTVNGPFAHKGCPNHLISYQGRIPYPRPGTCPPGGAFTPNMRTTKDFDDVVTFIR 402
QY 308 GDIDKVFRTS-SLK-G-----YHSSLNPRPGKCLPDQ-QP-I-PTETF--QVADRH 353
Db 403 NHPLMYNSTISPIHRRPLIVRIGTDYKTKIADRVNAADG-RYHVLFLGTDRTGTQKVVV 461
QY 354 EVAQRVEPMGPL-KTPLF-H--SKYHQVAVHRMQASHGETFHVLYLTDRGTHHKVVE 409
Db 462 LPTNSASGELILELEVFNKNHVPITTMETISSKKOOLYVSSNEGVSQVSLHRCHYGTAC 521
QY 410 -PGEHSAFNIQEPFRAAAIQTMSLDAERKLYVSSQWEYSQVPLDLCEYVGGGC 468
Db 522 ADCCLARDPYCAWDGHSRFRYPTGKRKRRODVRHGNPLTOCRGENLKAYRNAEIVQY 581
QY 469 HGCLMSRDPYCGWDGRCISIIYSSE--RSVLOSINPAEPHKECP--NPKDPKAPLOKVS 524
Db 582 GYRNSTFLECAPKPSOASIKWLLQDK 609
QY 525 A-PNSRYLSCPMESRHATYSWR-HKEN 550

RESULT 4 G01856 #type complete
ENTRY semaphorin V - human
TITLE #formal_name Homo sapiens #common_name man
ORGANISM 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
DATE 04-Sep-1998

ACCESSIONS	G01856
REFERENCE	G08634
#authors	Sekido, Y.
#submission	submitted to the EMBL Data Library, June 1995
#accession	G01856
##status	Preliminary; translated from GB/EMBL/DDBJ
##molecule_type	mRNA
##residues	1-749 ##label SEK
##cross-references	EMBL:U28369; NID:g974283; PID:g974284
CLASSIFICATION	#superfamily semaphorin
SUMMARY	#length 749 #molecular-weight 83121 #checksum 2747
Query Match	17.2% Score 716; DB 2; Length 749;
Best Local Similarity	33.0%; Pred. No. 5.42e-130;
Matches	167; Conservative 115; Mismatches 170; Indels 54; Gaps 41;
Db	112 ECNFKVLLHAYNRTHLLACGTCAFTCAFCVGHRAEPVLRLDPGRIGDEKGGSPYD 171 : :: : : : : : : : :
Qy	93 DCENITTLER-RSEGLLAGCTWARPSC-WNLV-N-G--TVVPL--G--EM-RGVAPES 141 : :: : : : : : : : :
Db	172 PRHRAASLVGEELYSGVA--ADLMGRDFTIFRSLGORPSLRTEPHDSRWLNPEPKFVKVEW 230 : :: : : : : : : : : : :
Qy	142 PDENSLVLFEGDEVYSTIRKEYNKR-IPFERIRGESELYT-S-DTV-MQNPFQIKAT- 196 : :: : : : : : : : : : :
Db	231 IPSENPNDDKIYFFPRETAWEAAPALGRUSVRGOICNDVGGQRST-VNKWTTFELKA 289 : :: : : : : : : : : : :
Qy	197 IVHQDAQYDDKIYFFREDNPDKNPE-APLNVSRVAQLCRGDQGESSLSVSKWNTEFLKA 255 : :: : : : : : : : : : :
Db	290 RLVCSPVGSGDTHFDLOLDVFLLS-SRDH-RTPLLIYAVFSTSSSIFOGSAYCVYSMDV 347 : :: : : : : : : : :
Qy	256 MLVCS--DAATNKNFNKLQDFLLPDPGQWRTRIVGYFSFPWN-Y--SAVCVYSLGDI 310 : :: : : : : : : : : : :
Db	348 RRAFLGPFAHKEGPMHOWSYQGRVPYPRPGMCPKSTFGTSFKDKFPDDVDIOFARNHPL 407 : :: : : : : : : : : : :
Qy	311 DK-V-F--RTSSL-K-CYHSSLNPRPGKCLPDQ-QPIP-TETF--QVAD--R-HPE 354 : :: : : : : : : : : : :
Db	408 MYNSVLEPTGG-R-PLFIQGANVTFTQIAADRVAAADGH-YDVLTGTDVGTVLKVISVP 464 : :: : : : : : : : : : :
Qy	355 VAORVEPMGPLKTPLF-H--SKHYOKVAVHRMQASHGETFHVLYLTDRGTJHKVVE-P 410 : :: : : : : : : : : : :
Db	465 KGRSPSAGLLLEHLHFVEDSAVTSMOISSKRHOLYVASRSAVAOTIALHRCRAHGVRT 524 : :: : : : : : : : : : :
Qy	411 -GQEHSFANIEIQFRRAAIIQMSLDAERKKLYVSQWEVSQVPLDLCEVYGGCH 469 : :: : : : : : : : : : :
Db	525 ECCLARPYCWDGACTRFPQSARRRFRQDVRNGDPTCLSCGDSRSSRALLEHKVFGVE 584 : :: : : : : : : : : : :
Qy	470 GCLMSRDPYCWDGDCRGCIYSIYSSERSVL--QSNPAEPKHECNPRPKRAPL-QKYSLAP 526 : :: : : : : : : : : : :
Db	585 GSASFLECEPRSLOARVENTFORAGV 610 : :: : : : : : : : : : :
Qy	527 NSRYILSCPMEISRHAITYSWR-HKENV 551 : :: : : : : : : : : : :
RESULT	5
ENTRY	G02173
TITLE	#type complete
ORGANISM	semaphorin iii family homolog - human
DATE	#formal_name Homo sapiens #common_name man 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 04-Sep-1998
ACCESSIONS	G02173
REFERENCE	G09275
#authors	Naylor, S.
#submission	submitted to the EMBL Data Library, October 1995
#accession	G02173
##status	Preliminary; translated from GB/EMBL/DDBJ
##molecule_type	mRNA
##residues	1-753 ##label NAY
##cross-references	EMBL:U38276; NID:g1061350; PID:g1061351
CLASSIFICATION	#superfamily semaphorin
SUMMARY	#length 753 #molecular-weight 84941 #checksum 5681
Query Match	16.5%; Score 688; DB 2; Length 753;
Best Local Similarity	31.314%; Pred. No. 1.50e-123;

```
QY 304 VYSLGDIKIDKVERTS-SLK-G--YH--S--S-LPNRPFGKCLPDQ-QPI-PTETF--QV-A 349
Db 402 FARSHPAMYNPFVINS-RPIMIKTIDVYQFTQIVWDRVDAEDGQ-YDVMFIGTDIGTVL 459
QY 350 -DR-HPEVAQRVEPMGLKPLFLHSK--YHYQKVAVHRMQASHGETFHVLYLTDRGTIH 405
Db 460 KWSIPKETWHELEEVLEEMTVREPTVISAMKISTKQOOLYIGSATYGSQPLPLHRCDV 519
QY 406 KVEE-PGEOEHSFAFIME-IOPFRAAAIQTMSLDAERKLYVSSQWESVQVPLDICEV 463
Db 520 YGKACAECLLARDPYCANDGSSCSRYPFTAKRTRRDIRNGDPLTHCSDLQHDNDHSGQ 579
QY 464 YGGCHGCLMSRDPYCGWDGRCISYSS-RSVL-QSINPAEPHKECPN-PRPKDAP-- 518
Db 580 TLEKIITYGVENSTFLECPKSORAIVY-WOFQKQNDHKEV 621
QY 519 -LQ-KVSLA-PNSRYILSC-PMESRHATYSWR-HKENVEQSCE 556

RESULT 7
ENTRY #158169 #type fragment
TITLE semaphorin III - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I58169
REFERENCE Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;
#authors resster-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#journal Neuron (1995) 14:949-959
#title Semaphorin III can function as a selective chemorepellent to
#cross-references MUID:95267432
#accession I58169
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-666 #label RES
#cross-references GB:L40484; NID:g703189; PID:g703190
GENETICS SemIII
#gene #superfamily semaphorin
CLASSIFICATION #length 666 #checksum 9654
SUMMARY
Query Match 16.0%; Score 667; DB 2; Length 666;
Best Local Similarity 30.1%; Pred. No. 9.84e-119;
Matches 150; Conservative 130; Mismatches 168; Indels 50; Gaps 39;

Db 60 RGKSPYDPKLLTASLLIDGELYSCTA-ANFMGRDFAIFRTLGHHPHPTREQHDSRWLNDP 118
QY 135 RGAPFSPDENSLVLFEGDEVYSTIRKQYNGK-IPFRRIERGESELYTS-DTV-MQNP 190
Db 119 RFISAHLIPESDNPEDDKVYFFEFRENAIDGHSKATHARIGQICKNDFGHRSL-VNWK 177
QY 191 QFIKATVHQ-DQAYDDKIYFFREDPNKPEAPLNVSRVAQLCRDQGESSLSVSKW 249
Db 178 TTFKLARLCSVPGPNIDTHDELQDVFML-NSKDP-KNPIYGVFTTSSNIFKGSVAVC 235
QY 250 NTFKLAMLVCS--DA-ATNKNFNRLQDVFLLPDPGQWRTRYGVFSNPNW-Y--SVC 303
Db 236 MYSMSDVRVLGPGYAHRDGPNYQWVYQGRVYPRPGTCPSKTFGGDSTKDLDDVIT 295
QY 304 VYSLGDIKDV---F--RTSS-LK--GYHSSLNPRPGKCLPDQ-QPI-PTETF--QV-A 349
Db 296 FARSHPAMYNPFVINS-RPIMIKTIDVYQFTQIVWDRVDAEDGQ-YDVMFIGTDIGTVL 353
QY 350 -DR-HPEVAQRVEPMGLKPLFLHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTIH 405
Db 354 KVSVPKETWHDLEEVLEEMTVREPTTISAMELSKQOOLYIGSTAGVAQLPLHRCDI 413
QY 406 KVEE-PGEOEHSFAFIME-IOPFRAAAIQTMSLDAERKLYVSSQWESVQVPLDICEV 463
Db 414 YGKACAECLLARDPYCANDGSSCSRYPFTAKRTRRDIRNGDPLTHCSDLQHDNDHHPG 473
```

```
QY 464 YGGCHGCLMSRDPYCGWDGRCISYSS-RSVL-QSINPAEPHKECPN-PRPKD--AP 518
Db 474 SLERIITYGVENSTFLECPKSORALVYQFORNEDRKEEIKMGDHIIRTEQGLLLRS 533
QY 519 -LQ-KVSLA-PNSRYILSC-PMESRHATY-SW-RHRENVEQSCEPG-HQ-SPNCILLFIEN 570
Db 534 LOKKDSGNYLCHAVEHGF 551
QY 571 LTAQYGHYFCEAQEGSY 588

RESULT 8
ENTRY #148747 #type complete
TITLE semaphorin D - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48747
REFERENCE Puschel, A.W.; Adams, R.H.; Betz, H.
#authors Neuron (1995) 14:941-948
#journal Murine semaphorin D/collapsin is a member of a diverse gene
#title family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48747
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-772 #label RES
#cross-references EMBL:X85993; NID:g854329; PID:g854330
GENETICS semD
#gene #superfamily semaphorin
CLASSIFICATION #length 772 #molecular-weight 88710 #checksum 1776
SUMMARY
Query Match 16.0%; Score 666; DB 2; Length 772;
Best Local Similarity 31.0%; Pred. No. 1.67e-118;
Matches 140; Conservative 119; Mismatches 146; Indels 47; Gaps 38;

Db 166 RGKSPYDPKLLTASLLIDGELYSCTA-ADFMGRDFAIFRTLGHHPHPTREQHDSRWLNDP 224
QY 135 RGAPFSPDENSLVLFEGDEVYSTIRKQYNGK-IPFRRIERGESELYTS--DIV-MQNP 190
Db 225 RFISAHLIPESDNPEDDKVYFFEFRENAIDGHSKATHARIGQICKNDFGHRSL-VNWK 283
QY 191 QFIKATVHQ-DQAYDDKIYFFREDPNKPEAPLNVSRVAQLCRDQGESSLSVSKW 249
Db 284 TTFKLARLCSVPGPNIDTHDELQDVFML-NSKDP-KNPIYGVFTTSSNIFKGSVAVC 341
QY 250 NTFKLAMLVCS--DA-ATNKNFNRLQDVFLLPDPGQWRTRYGVFSNPNW-Y--SVC 303
Db 342 MYSMSDVRVLGPGYAHRDGPNYQWVYQGRVYPRPGTCPSKTFGGDSTKDLDDVIT 401
QY 304 VYSLGDIKDVRTS-SLK-G--YH--S--S-LPNRPFGKCLPDQ-QPI-PTETF--QV-A 349
Db 402 FGRSHPAMYNPFVINS-RPIMIKTIDVYQFTQIVWDRVDAEDGQ-YDVMFIGTDIGTVL 459
QY 350 -DR-HPEVAQRVEPMGLKPLFLHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTIH 405
Db 460 KVSVPKETWHDLEEVLEEMTVREPTTISAMELSKQOOLYIGSTAGVAQLPLHRCDI 519
QY 406 KVEE-PGEOEHSFAFIME-IOPFRAAAIQTMSLDAERKLYVSSQWESVQVPLDICEV 463
Db 520 YGKACAECLLARDPYCANDGSSCSRYPFTAKRTRRDIRNGDPLTHCSDLQHDNDHHPG 579
QY 464 YGGCHGCLMSRDPYCGWDGRCISYSS-RSVL-QSINPAEPHKECPN-PRPKD--AP 518
Db 580 SLERIITYGVENSTFLECPKSORALVY-WQ 610
QY 519 -LQ-KVSLA-PNSRYILSC-PMESRHATYSWR 546

RESULT 9
ENTRY #D49423 #type complete
```


TITLE semaphorin III precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 04-Sep-1998

ACCESSIONS D49423
REFERENCE A49423
#authors Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
#journal Cell (1993) 75:1389-1399
#title The Semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules.

#accession D49423
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-771 #label KOL
##cross-references GB:L26081; NID:g799328; PID:g436560

GENETICS
#gene GDB:SEMA1
#cross-references GDB:283448
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 771 #molecular-weight 88889 #checksum 6249

Query Match 15.9%; Score 661; DB 2; Length 771;
Best Local Similarity 30.3%; Pred. No. 2.33e-117;
Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;

Db 166 RGKSPDPKLLTASLIDGELYSSTA-ADFMGRDFAIFRTLGHHHHPIRTEQHDNRWLDNP 224
QY 135 RGYAPSPDENSLSVLDEGEDEVYTIKKQYNGK-IPFRRIIRGESELYTS--DTV-MQNP 190
Db 225 KFTSAHLSIEDNPEDKVVYFFERNAIDGESHGKATHARIGQICNDGEGHSL-VNWK 283
QY 191 QFIKATIVHQ-DOAYDDKTYIFFREDNPKNPEAPLNVSRVAQLCGDGGESSLSVSKW 249
Db 284 TTFKLKARLCSVPGPGNGIDTHEDELQDFELM-NFKDP-KNPVYVGYFTTSSNIFKGSVC 341
QY 250 NTFKLKAMLVCS--DA-ATNKNENRLQDVELLPDPGQWRDTRVYGVFSPWN-Y--SVC 303
Db 342 MYSMSVRRVFLGPAHROGNQWVPYQGRVPPRGTCPSFTFGGFSTKDLDPDVT 401
QY 304 VYSLGIDKLVFRTS-SLK-G-YH--S--S-LPNRPGRKLPDQ-QPI-PTETF--QV-A 349
Db 402 FARSHPAMYNPVPMMNRPIVIKTDVNYQFTQIVDVRDAEDGO-YDVMFIDTGVTLK 460
QY 350 -DR-HPEVAQRVPEMGLTKPL-FHSKYHQKVAHRMQASHGETFHVLYLTDRGTIRK 406
Db 461 VVSIPKETWYDLLEVLLEMTVFREPTAISAMELSKQOOLYIGCTAGVAQLPLHRCDIY 520
QY 407 VVE-PGEOEHSFAFNME-IQPFRAAAIQTMSLDAERRKLYVSSQWVSOVPLDICEVY 464
Db 521 GKACAECLLARDPYCAWDGACSRYPFTAKRRTTRDIRNGDPLTHCSDLHHDNHHGSP 580
QY 465 GGGCHGCLMSRDPYCGWDQGRCSISSE-RSVL-QSINPAEPHKECPNPKDK---AP 518
Db 581 EERIIYGVENSSTFLCSPKSRQALVYQFORNEERKEIRVDDHIIRTDQGLLRSLO 640
QY 519 LQKVSUA-PNSRYLSC-PMESRHATY-SW-RHKENY-EQSCPEPGHQ-SPNCILFTENLT 572
Db 641 QKDSGNVLCHAVEHGFIO 658
QY 573 AQQYGHYFCEAQEGSYFR 590

RESULT 10
ENTRY I48746 #type fragment
TITLE semaphorin C - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 28-Feb-1997

ACCESSIONS I48746
REFERENCE I48744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene

#cross-references MUID:95267431
#accession I48746
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-782 #label RES
##cross-references EMBL:X85992; NID:g854327; PID:g854328

GENETICS
#gene semC
SUMMARY #length 782 #checksum 1571

Query Match 14.8%; Score 618; DB 2; Length 782;
Best Local Similarity 33.7%; Pred. No. 1.55e-107;
Matches 149; Conservative 96; Mismatches 148; Indels 49; Gaps 32;

Db 74 KRDCQNYIKILLPLNSHLLTCGTAAFSPLCAYIHIAFTLAQDEAGNVILEDGKHCPEF 133
QY 91 KRDCENYIT-LLERRSEGLLACGTNARHPSC-W-N-----LV-NGTV-VPLGEMRGYAPF 140
Db 134 DPNFKSTALVVDGELYTGTVS-SFGQNDPAISRSQSRRPTKTESSLNLWLODPAFVASATS 192
QY 141 SPDENSILVLFEGDEVYSTIRKQFYNGKIPFRRIIRGESELYTSDTV--MQNPQFIKATIV 198
Db 193 PSLSGSPIGDDDKIYFFFESETGOEPEFFENTI-VSRVARVCKGDEGERVLQ-QRWTSEL 250
QY 199 HQDQ-A-Y--DKIIYFFREDNPD-KNPEAPLNVSRVAOLCRQDGGESSLSVSKWNTFL 253
Db 251 KAQLCSRPDDGFPFNVLDQVFTLNPNQ-DWRKTLISIGVFTSQWHRGTTEGSAICVFTM 309
QY 254 KAMLVCSDAATNKNFNRLQDVELL-PDPSQWRDTRVYGVFSPWN---Y--SACVYSL 307
Db 310 NDVQKAFDGLYKKNRETQWTTTETHQVTPRPACITNSAREKINSLSQLPDRVLNLF 369
QY 308 GDIDKVF-----RTSS-LKG-Y-HSS-LPNRPGRKLPD--QOPIPTETQVADRHEVA 356
Db 370 KOHFLMDGOVRSLLLQPRARYQVRVAVHRVCLHS-TYDVLFLGTDGGLRHKAVTLSSR 428
QY 357 QRVEPM-GLPKTLF--HSKYHQKVAHRMQASHGETFHVLYLTDRGTIRHVKVEPGEQ 413
Db 429 VH---I-IEELQIFPOGQPVQVNLDSLHGLLYASHSGVQVVPVANCSLYPT-CGDCLL 483
QY 414 EHSFAFNIMEIQPFRAAAIQTMSLDAERRKLYVSSQWVSOVPLDICEVYGGGCHGCLM 473
Db 484 ARDPYCAWTGSACRLASLYOPD 505
QY 474 SRDPYCGWDQGRG-I-SIYSS 493

RESULT 11
ENTRY S66498 #type complete
TITLE M-sema F protein precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 10-Sep-1997

ACCESSIONS S66498
REFERENCE S66498
#authors Inagaki, S.; Furuyama, T.; Iwahashi, Y.
#journal FEBS Lett. (1995) 370:269-272
#title Identification of a member of mouse semaphorin family.
#cross-references MUID:95385809
#accession S66498
#status preliminary
#molecule_type mRNA
#residues 1-834 #label INA
##cross-references EMBL:S79463; NID:g1110598; PID:g1110599

FEATURE
1-21 #domain signal sequence #status predicted #label SIG\
22-834 #product M-sema F protein #status predicted #label MAT
SUMMARY #length 834 #molecular-weight 92556 #checksum 7189

Query Match 12.5%; Score 522; DB 2; Length 834;
Best Local Similarity 28.5%; Pred. No. 7.96e-86;
Matches 142; Conservative 118; Mismatches 190; Indels 49; Gaps 36;

ORGANISM #formal_name vaccinia virus
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997

ACCESSIONS S29921
REFERENCE S29907
#authors Amegadzie, B.Y.
#submission submitted to the EMBL Data Library, January 1991
#accession S29921
#status preliminary
#molecule_type DNA
#residues 1-441 #label AME
#cross-references EMBL:X57318; NID:g62239; PID:g62254
SUMMARY #length 441 #molecular_weight 50185 #checksum 6034

Query Match 11.1%; Score 463; DB 2; Length 441;
Best Local Similarity 33.9%; Pred. No. 1,14e-72;
Matches 87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;

Db 115 LVCGTNGNPKCWK-IDGSDDPKRGGRGYAPYQNSKVITIISHNGC-VLSDINISK-EG-I 170
QY 109 LACGTNARHPSCWNLVNGTVPLGEMRGYAPFSPDENSELVLFEGDEVYSTIRKQYNGKI 168
Db 171 KWRFRDPCGCDLYTADNVIPKDG-LRGAFAVDKGTYP-KVILFTDITIGSKR--I-VK 225
QY 169 PRFRIRGE-S-ELTSDVMQNPQFIKATIVHQDAYDDKIYFFREDNPKNPEAPLN 226
Db 226 IPYIAQMLNDGGPSSLSHRKSTFLKVELEC-DID-GRSY-R-Q-IIHSRTIKTD-ND 279
QY 227 VSRVAQLCRGDGESSLSVSKWNTFLKAMLVCSDAATNKNRLQDVLLPDPSSQWRD 286
Db 280 TILYVFFDPYSKALCTYSMTIKOSFSTKLEGYTKLPSPASGICLPAGKYVPHHTF 339
QY 287 TRVYGVSFNWYSACVYVSLGDIDKVFRTSSLSKGLHSLPNRPCKLDPDQOPIPTET 346
Db 340 EVIEKYNVLDDIIKPLS 356
QY 347 QVADRHPEVAQRVEPG 363

RESULT 15
ENTRY JH0798 #type complete
TITLE fasciclin IV precursor - American bird grasshopper
ORGANISM #formal_name Schistocerca americana #common_name American bird grasshopper
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Sep-1997
ACCESSIONS JH0798
REFERENCE JH0798
#authors Kolodkin, A.; Matthes, D.J.; O'Connor, T.P.; Patel, N.H.; Admon, A.; Bentley, D.; Goodman, C.S.
#journal Neuron (1992) 9:831-845
#title Fasciclin IV: Sequence, expression, and function during growth cone guidance in the grasshopper embryo.
#accession JH0798
#molecule_type mRNA
#residues 1-730 #label KOL
#cross-references GB:L00709; NID:g160844; PID:g160845
#experimental_source embryo
COMMENT This protein plays a role in growth cone guidance in the developing central nervous system.
glycoprotein; transmembrane protein

KEYWORDS
FEATURE
1-22 #domain signal sequence #status predicted #label SIG\
23-730 #product fasciclin IV #status predicted #label MAT\
23-627 #domain extracellular #status predicted #label EXT\
628-652 #domain transmembrane #status predicted #label TMW\
653-730 #domain intracellular #status predicted #label INT\
44,71,163,267,360, #binding_site carbohydrate (Asn) (covalent) #status predicted
539

SUMMARY #length 730 #molecular_weight 81214 #checksum 5881
Query Match 10.7%; Score 447; DB 2; Length 730;

Best Local Similarity 28.4%; Pred. No. 3.94e-69;
Matches 124; Conservative 114; Mismatches 153; Indels 46; Gaps 38;

Db 106 DCQNTIRVLAKIDDDRLVICGTNAVKPLCRHVALKDGDDVVEKEVEGRGLCPDPDHNST 165
QY 93 DCENITILLERSEG-LLACGTNARHPSCWN--LYNGT-VVPLG-EMRGYAPFSPDENSEL 147
Db 166 AIYSEGQILYSATV-ADFSCTDP-LI-YRG-P-LRTERS-DL-K-QLNAPNFVN-TMEYNDF 217
QY 148 VLFEGDEVYSTIRKQYNGKIYFFREDNPDKIPFRIRGESELYTSDTVMQNPQFIKATIVHQDAYDDK 207
Db 218 IFFFFRETAVEYINGKAIY-SRVARVCKHDKGGPHOFG-DRWTSFLKSLRLNCSVPDGYD 275
QY 208 IYFFREDNPDKIPFRIRGESELYTSDTVMQNPQFIKATIVHQDAYDDK 266
Db 276 FYFNEIQSTSDIIEGNYGGO-VEKLIYGVFTTPVNSIGGSACVAFSMKSIILESFDPGPEK 334
QY 267 -NENLQDVF-LLPDP-SQWRDTRVYGVFSNPNW-Y--SACVYSL-G--D-IDKVFRT 316
Db 335 QETMNSNMLAVPSLKVPEPRPGQCVNDSRTLDPVSVNFVKSHSTLMDAEPAFET-RPILI 393
QY 317 -SSLKG-YHS--SL--PNPRPGKCLPDQOPIPTETFOVADRHPEVAQRVEPNGLKTPLF 370
Db 394 RISLQYRFTKIAVDQOVRTPDGKAYDVLFIGTDDGKVIKALNSAFSDSDTVDVSVVIEL 453
QY 371 H-S-KYHYQKAV-HRMQASHGETFHVLYLTDRGTIHKVVEPG--EQEHSFAFNIME-I 424
Db 454 QVLPGPVPKLYVVRMDGDDSKLVVSDDEILAKLHRCGSDKITNCRECVSLQDPPYCA 513
QY 425 QPFRRAAIIQTY--SLDAERKKLYVSSQWEVSQVPLDLCEVYG-GGCHGCLMSRDPYCG 480
Db 514 WDNVELKCTAVGSPDWS 530
QY 481 WQD-G-RCISIIYSERS 495

Search completed: Thu Jul 8 19:28:15 1999
Job time : 32 secs.

THIS PAGE BLANK (USPTO)

WQELH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:26:01 1999; Maspar time 17.59 Seconds
Tabular output not generated. 898.289 Million cell updates/sec

Title: >US-09-041-236-2
Description: (48-606) from US09041236.pep (35 of 45)
Perfect Score: 4170
Sequence: 1 HEPGSSVWVGGRKVLFD.....SYFRAQHMLLPEDGIMAE 559

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: swiss-prot37
1:swissprot

Statistics: Mean 49.993; Variance 75.468; scale 0.662

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	470	11.3	403	1	VA39_VACC	1.59e-86
2	463	11.1	441	1	VA39_VACCV	1.02e-84
3	163	3.9	1871	1	SEX_HUMAN	7.18e-13
4	105	2.5	227	1	US08_HCMVA	3.29e-02
5	102	2.4	275	1	NRT2_RAT	9.69e-02
6	99	2.4	275	1	NT11_RAT	2.78e-01
7	99	2.4	460	1	EF1A_TTIRE	2.78e-01
8	99	2.4	591	1	VRP2_SALTY	2.78e-01
9	99	2.4	591	1	VRP2_SALCH	2.78e-01
10	99	2.4	591	1	VRP2_SALEN	2.78e-01
11	99	2.4	593	1	VRP2_SALDU	2.78e-01
12	99	2.4	916	1	PM1A_AJECA	2.78e-01
13	98	2.4	1132	1	DNBI_HSV60	3.92e-01
14	95	2.3	255	1	HA21_HUMAN	1.09e+00
15	94	2.3	329	1	RBSR_ECOLI	1.51e+00
16	95	2.3	518	1	VL2_HPV5B	1.09e+00
17	95	2.3	518	1	VL2_HPV05	1.09e+00
18	95	2.3	683	1	AMOR_ARTGO	1.09e+00
19	94	2.3	947	1	PM22_YEAST	1.51e+00
20	94	2.3	982	1	POL_HTLV2	1.51e+00
21	94	2.3	1069	1	ENTK_MOUSE	1.51e+00
22	94	2.3	1115	1	IREL_YEAST	1.51e+00
23	94	2.3	1200	1	DDX8_CAEEL	1.51e+00

24	95	2.3	1390	1	MET_HUMAN	HEPATOCYTE GROWTH FACT	1.09e+00
25	90	2.2	119	1	RL19_MYCPN	50S RIBOSOMAL PROTEIN	5.53e+00
26	90	2.2	213	1	AMEX_BOVIN	AMELOGENIN, CLASS I PR	5.53e+00
27	90	2.2	254	1	HA22_HUMAN	HLA CLASS II HISTOCOMP	5.53e+00
28	91	2.2	348	1	DDL_ENTFA	D-ALANINE--D-ALANINE L	4.02e+00
29	91	2.2	356	1	VP39_NPVLD	MAJOR CAPSID PROTEIN.	4.02e+00
30	93	2.2	380	1	DP3B_MYCPN	DNA POLYMERASE III, BE	2.10e+00
31	92	2.2	442	1	CEMI_YEAST	3-OXOACYL-L-ACYL-CARRIE	2.91e+00
32	92	2.2	460	1	EF1A_NEUCR	ELONGATION FACTOR 1-AL	2.91e+00
33	91	2.2	498	1	ELAS_PSEAE	PSEUDOLYSIN PRECURSOR	4.02e+00
34	91	2.2	598	1	NOT_HUMAN	IMMEDIATE-EARLY RESPON	4.02e+00
35	91	2.2	707	1	GCVK_HCMVA	GANCICLOVIR KINASE (EC	4.02e+00
36	93	2.2	1029	1	ENDL_YEAST	VACUOLAR BIOGENESIS PR	2.10e+00
37	92	2.2	1132	1	DNBI_HSV62	MAJOR DNA-BINDING PROT	2.91e+00
38	92	2.2	1302	1	MDR5_DROME	MULTIDRUG RESISTANCE P	2.91e+00
39	92	2.2	1663	1	CO3_MOUSE	COMPLEMENT C3 PRECURSO	2.91e+00
40	91	2.2	1666	1	CO3_CAVPO	COMPLEMENT C3 PRECURSO	4.02e+00
41	90	2.2	1874	1	POLR_KYMWJ	RNA REPLICASE POLYPROT	5.53e+00
42	89	2.1	404	1	NIFS_ECOLI	NIFS PROTEIN HOMOLOG.	7.58e+00
43	89	2.1	437	1	RFBF_MYXXA	O-ANTIGEN EXPORT SYSTE	7.58e+00
44	89	2.1	598	1	RNRI_RAT	REGENERATING LIVER NUC	7.58e+00
45	89	2.1	649	1	RA32_SCHPO	DNA REPAIR PROTEIN RAD	7.58e+00

ALIGNMENTS

RESULT	1	STANDARD	PRT	403 AA.
ID	VA39_VACC			
AC	P21062;			
DT	01-FEB-1991 (REL. 17, CREATED)			
DT	01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)			
DE	PROTEIN A39.			
GN	A39R.			
OS	VACCINIA VIRUS (STRAIN COPENHAGEN).			
OC	VIRUSES: DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;			
OC	ORTHOPOXVIRUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 91021027.			
RA	GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,			
RA	PAOLETTI E.;			
RT	"The complete DNA sequence of vaccinia virus.";			
RL	VIROLOGY 179:247-266(1990).			
RN	[2]			
RP	COMPLETE GENOME.			
RA	GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,			
RA	PAOLETTI E.;			
RL	VIROLOGY 179:517-563(1990).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M35027; G335517; -			
DR	PIR; E42521; E42521.			
SQ	SEQUENCE 403 AA; 45741 MW; FE0DB9AB CRC32;			

Query Match 11.3%; Score 470; DB 1; Length 403;				
Best Local Similarity 33.5%; Pred. No. 1.59e-86;				
Matches 86; Conservative 59; Mismatches 95; Indels 17; Gaps 16;				

Db	77	LVCGNNPNKCKW-IDGSDDPKHRGRGYAPQNSKVITISYN-ECVLSDINISK-EG-I	132
Qy	109	LACGTNHRHPSCNLVNGTVVPLGEMRGYAPFSPDENSELVLFEGDEVSTIRKQEVNGKI	168
Db	133	KWRREFDPCGYDLYTADNVTPKDG-LRGAFVDKDGTYD-KVYILFTDTIGSKR--I-VK	187
Qy	169	PRFRIRGE-S-ELTSDTVMNQPOFIKATIVHQDAYDDKIYFFREDNPDKNPEAPLN	226

13-JUL-1990 (REL. 36, LAST ANNOTATION UPDATE)
T-CELL ESTRO-ADP-RIBOSYLTRANSFERASE 2 PRECURSOR (EC 2.4.2.31) (T-CELL
DE NAD(P)(+)-ARGININE ADP-RIBOSYLTRANSFERASE 2) (T-CELL MONO(ADP-
DE RIBOSYL)TRANSFERASE 2) (ALLOANTIGEN RT6.2) (T-CELL SURFACE PROTEIN
DE RT6.2).
GN ART2B OR RT6-B

db 194 IGVVTKFEES-FRBDQEE-VITPGVEVWOKVPTOCYNFTETDSQKPKKSNVNCIYSSACAP 251

[illegible]

QY 131 LGE-MRGYAPFSPDENSILVLFEGDEVYSTIRKOEYNGK-IPRFRIRGESE-LYTSDTVM 187
Db 252 ES 253
QY 188 QN 189

RESULT 6
ID NRTL_RAT STANDARD; PRT; 275 AA.
AC P17982;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 (T-CELL MONO(ADP-
DE NAD(P)(+)-L-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL MONO(ADP-
DE RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN
DE RT6.1).
GN ART2A OR RT6-A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEWIS A;
RX MEDLINE; 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell
RT allouantigen RT6.1";
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
RN [2]
RP MUTAGENESIS OF GLN-207.
RX MEDLINE; 96275529.
RA MAEHAMA T., HOSHINO S.-I., KATADA T.;
RT "Increase in adp-ribosyltransferase activity of rat T lymphocyte
RT allouantigen RT6.1 by a single amino acid mutation.";
RL FEBS LETT. 388:189-191(1996).
CC -!- FUNCTION: HAS NAD+ GLYCOPOLYMERASE ACTIVITY AND EXTREMELY LOW ADP-
CC RIBOSYLTRANSFERASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
CC N2-(ADP-D-RIBOSYL)-L-ARGININE
CC -!- CATALYTIC ACTIVITY: NAD(+) + H(2)O = NICOTINAMIDE + ADP-RIBOSE.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -!- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52082; G57168; -.
CC EMBL; M31138; G206804; -.
CC PIR; S08464; S08464.
CC PROSITE; PS01291; ART; 1.
CC PFAM; PF01129; ART; 1.
CC TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
CC T-CELL DIFFERENTIATION; GPI-ANCHOR.
KW
FT SIGNAL 1 20 T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
FT CHAIN 21 246 HYDROPHOBIC, REMOVED DURING MATURATION
FT PROPEP 247 275 (BY SIMILARITY).
FT GPI-ANCHOR (BY SIMILARITY).
FT LIPID 246 246 BY SIMILARITY.
FT ACT_SITE 209 209 POTENTIAL.
FT CARBOHYD 58 58 Q->E: INCREASED ADP-RIBOSYLTRANSFERASE
FT MUTAGEN 207 207 ACTIVITY.
FT
SQ SEQUENCE 275 AA; 31388 MW; 523B1A84 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 2.78e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;

Db 194 LGVYIKERS-FYPDQEE-VLIPGYEVYQVKTQGYNEIFLDSFKRKKSNYNCLYSAGTR 251
QY 131 LGE-MRGYAPFSPDENSILVLFEGDEVYSTIRKOEYNGK-IPRFRIRGESE-LYTSDTVM 187
Db 252 ES 253
QY 188 QN 189

RESULT 7
ID EFIA_TRIRE STANDARD; PRT; 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEFL.
OS TRICHODERMA REESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-QM9414 / RUTC-30;
RX MEDLINE; 94124021.
RA NAKARI T., ALATALO E., PENTTILA M.;
RT "Isolation of Trichoderma reesei genes highly expressed on glucose-
RT containing media: characterization of the tefl gene encoding
RT translation elongation factor 1 alpha";
RL GENE 136:313-318(1993).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-tRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z23012; G312887; -.
CC PIR; S35772; S35772.
CC PROSITE; PS00301; EFACITOR_GTP; 1.
CC PFAM; PF00009; GTP_EFTU; 1.
CC HSP; P07157; IAIIP.
CC ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
KW NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 49830 MW; B9ABABB2 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 2.78e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;

Db 210 YKGWEKEIKAGFTG-KTLLLEAIDSIERPGR-PTDKPLRLPQDV 252
QY 478 YCGWDGGRCSISYSSERSVLQSIINPAEPHKCPNPKPKAPLQKV 522

RESULT 8
ID VRP2_SALTY STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)


```

RN SEQUENCE FROM N.A.
RX MEDLINE: 95266321.
RA GOMPEL U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution.";
RL VIROLOGY 209:29-51(1995).
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES' DNA-BINDING PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X92436; G1044871; -
CC EMBL: X83413; G854020; -
CC PFAM: PF00747; Viral_DNA_bp; 1.
CC DNA-BINDING; DNA REPLICATION; ZINC-FINGER; NUCLEAR PROTEIN.
CC ZN_FING 459 475 C4-TYPE.
CC SEQUENCE 1132 AA; 127762 MW; DB699A02 CRC32;
Query Match 2.4%; Score 98; DB 1; Length 1132;
Best Local Similarity 28.2%; Pred. No. 3.92e-01;
Matches 24; Conservative 18; Mismatches 37; Indels 6; Gaps 6;
Db 435 WNLNRMVYVGNAGNHAETI-YNHLVNCNANLCEFCGKC-C-QSCIGTAMVRVGRRLPAIP 491
QY 297 WNYGAVCYVSLGIDKVFRTSSLGXHYHSLNPRGKLPDQQPPTFTFOVADRHPEVA 356
Db 492 KNVK-KEPLVMSMF-SRY-YAEVDI 513
QY 357 QRVEPMGLKPLFLHRSKHYHKVAV 381
RESULT 14
ID HA21-HUMAN STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN
RP SEQUENCE FROM N.A.
RX MEDLINE: 84168117.
RA AUFFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPES D.,
RA STROMINGER J.L.;
RT "Isotypic and allotypic variation of human class II
RT histocompatibility antigen alpha-chain genes.";
RL NATURE 308:327-333(1984).
RN
RP SEQUENCE OF 40-255 FROM N.A.
RX MEDLINE: 83065171.
RA AUFFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT "cDNA clone for the heavy chain of the human B cell alloantigen DC1:
RT strong sequence homology to the HLA-DR heavy chain.";
RL PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
RN
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J00199; -; NOT_ANNOTATED_CDS.
CC PIR: A02212; HLHUCD.
CC PROSITE: PS00290; IG_MHC; 1.
CC PFAM: PF00047; Ig; 1.
CC PFAM: PF00993; MHC_II_alpha; 1.
CC HSP; P01910; IIAK.
CC MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
KW SIGNAL 1 23
FT CHAIN 24 255 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT DQ(1) ALPHA CHAIN.
FT DOMAIN 24 110 EXTRACELLULAR ALPHA-1.
FT DOMAIN 111 204 EXTRACELLULAR ALPHA-2.
FT DOMAIN 205 217 CONNECTING PEPTIDE.
FT TRANSMEM 218 240
FT DOMAIN 241 255 CYTOPLASMIC TAIL.
FT DISULFID 133 189 BY SIMILARITY.
FT CARBOHYD 104 104 POTENTIAL.
FT CARBOHYD 144 144
SQ SEQUENCE 255 AA; 28105 MW; 5C46DD59 CRC32;
Query Match 2.3%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1.09e+00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;
Db 25 DIVADSVAAQLG-VNLYQSYGPGSGYSGHEFDGDEEYVDLERKETVWQLPFRFR 78
QY 122 NLVNGTVPLGEMGYAPSPDENSELVLFEGDEV-YSTIRKQYNGKIPRRIR 175
RESULT 15
ID RBSR_ECOLI STANDARD; PRT; 329 AA.
AC P25551;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOSE OPERON REPRESSOR.
DE RBSR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6 AND 250-256.
RX STRAIN-K12;
RX MEDLINE: 93278299.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural and functional analyses of the repressor, RbsR, of the
RT ribose operon of Escherichia coli.";
RL PROTEIN SCI. 1:831-842(1992).
RN
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MGI655;
RX MEDLINE: 93315143.
RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL GENOMICS 16:551-561(1993).
RN
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RA IIDA A., TESHIBA S., MIZOBUCHI K.;
RL SUBMITTED (XXX-1993); TO EMBL/GENBANK/DBDJ DATA BANKS.
RN
RP SIMILARITY TO RIBOSE-BINDING PROTEINS.
RX MEDLINE: 93278300.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural homology between rbs repressor and ribose binding protein
RT implies functional similarity.";
RL PROTEIN SCI. 1:843-849(1992).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RBSDACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING

```

W P E R E H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:26:39 1999; Maspar time 37.13 Seconds
821.600 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (48-606) from US09041236.pep (35 of 45)
Perfect Score: 4170
Sequence: 1 HEPGSSVWVGGRGKYLFD.....SYFRAQHQLLPEDGIMAE 559

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrmb19
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 48.497; Variance 75.227; scale 0.645

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description Pred. No.
1	4170	100.0	666 4	075326 SEMAPHORIN L. 0.00e+00
2	2165	51.9	393 11	088371 SEMAPHORIN L (FRAGMENT) 0.00e+00
3	1771	42.5	653 14	084906 SIMILAR TO GENBANK ACC 0.00e+00
4	737	17.7	748 11	Q62177 SEMAPHORIN A PRECURSOR 1.36e-153
5	725	17.4	751 11	Q62181 SEMAPHORIN E PRECURSOR 1.89e-150
6	716	17.2	749 4	Q13214 SEMAPHORIN V. 4.29e-148
7	716	17.2	750 4	Q93018 SEMAPHORIN V. 4.29e-148
8	710	17.0	751 4	Q92985 SEMAPHORIN E. 1.59e-146
9	705	16.9	751 13	042236 COLLAPSIN 3. 3.23e-145
10	688	16.5	753 4	Q13372 SEMAPHORIN III FAMILY 8.89e-141
11	682	16.4	754 11	088633 SEMAPHORIN IV ISOFORM 3.27e-139
12	682	16.4	785 11	088632 SEMAPHORIN IV ISOFORM 3.27e-139
13	681	16.3	772 13	090607 COLLAPSIN. 5.95e-139
14	681	16.3	785 4	Q13275 SEMAPHORIN IV. 5.95e-139
15	679	16.3	785 4	Q13704 SEMAPHORIN. 1.98e-138
16	668	16.0	772 11	Q63548 SEMAPHORIN III/COLLAPS 1.45e-135
17	666	16.0	772 11	Q62180 SEMAPHORIN D PRECURSOR 4.81e-135
18	661	15.9	771 4	Q14563 SEMAPHORIN-III. 9.63e-134
19	618	14.8	782 11	Q62179 SEMAPHORIN C (SEM C) (1.38e-122
20	614	14.7	775 11	P70275 SEMAPHORIN H. 1.50e-121

21	599	14.4	861 11	009126 SEMAPHORIN J (SEMAPHOR	1.13e-117
22	594	14.2	761 13	Q90663 COLLAPSIN-2.	2.19e-116
23	588	14.1	775 4	015041 KIAA0331.	7.73e-115
24	572	13.7	785 13	042237 COLLAPSIN 5.	1.01e-110
25	568	13.6	294 13	Q90664 COLLAPSIN-3 (FRAGMENT)	1.08e-109
26	544	13.0	862 4	Q92854 SEMAPHORIN.	1.54e-103
27	522	12.5	834 11	Q64151 SEMAPHORIN I (M-SEMA F	6.41e-98
28	511	12.3	299 13	Q90666 COLLAPSIN-5 (FRAGMENT)	4.05e-95
29	480	11.5	760 11	Q62178 SEMAPHORIN B PRECURSOR	2.93e-87
30	473	11.3	295 13	Q90665 COLLAPSIN-4 (FRAGMENT)	1.72e-85
31	447	10.7	730 5	Q26473 FASCICLIN IV.	5.94e-79
32	448	10.7	1074 4	Q13591 SEMAPHORIN F HOMOLOG.	3.33e-79
33	441	10.6	1077 11	Q62217 SEMAPHORIN F PRECURSOR	1.89e-77
34	416	10.0	1093 11	Q60519 SEMAPHORIN G PRECURSOR	3.25e-71
35	390	9.4	494 4	060408 SEMAPHORIN F (FRAGMENT	8.95e-65
36	392	9.4	888 11	Q35464 SEMAPHORIN VIA.	2.84e-65
37	374	9.0	712 5	Q26972 SEMAPHORIN-I PRECURSOR	7.56e-61
38	367	8.8	562 5	Q17330 CESEMA.	3.90e-59
39	367	8.8	771 5	Q24322 SEMAPHORIN-I.	3.90e-59
40	362	8.7	706 5	Q24323 SEMAPHORIN-II.	6.48e-58
41	341	8.2	284 11	Q34948 SEMAPHORIN IV HOMOLOG	8.12e-53
42	316	7.6	886 11	O54951 SEMAPHORIN N (SEMAPHOR	8.20e-47
43	316	7.6	887 11	O70141 SEMAPHORIN Z.	5.20e-47
44	283	6.8	770 5	O44253 SIMILAR TO SEMAPHORIN-	5.16e-39
45	204	4.9	1963 4	O75051 KIAA0463 PROTEIN (FRAG	3.65e-21

ALIGNMENTS

RESULT 1
ID 075326 PRELIMINARY; PRT; 666 AA.
AC 075326;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN L.
GN SEMAL
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses".
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030698; G3523115; -;
SQ SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;

Query Match				
Best Local Similarity 100.0%; Score 4170; DB 4; Length 666;				
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Db	80	HEPGSSVWVGGRGKYLFD	PEPGKNASVTVNTGSKGCLDRDCENYITLLRRSEG	139
Qy	48	HEPGSSVWVGGRGKYLFD	PEPGKNASVTVNTGSKGCLDRDCENYITLLRRSEG	107
Db	140	LLAGCTNARHPSCNVLNVTVP	LGEMRGYAPFSPDENSLVLFEGDEVYSTIRKOEYNGK	199
Qy	108	LLAGCTNARHPSCNVLNVTVP	LGEMRGYAPFSPDENSLVLFEGDEVYSTIRKOEYNGK	167
Db	200	IPRFRIRGESELYTSDTVMQNPQ	FIKATIVHQDQAYDDKIYFFREDNPKNPEAPLN	259
Qy	168	IPRFRIRGESELYTSDTVMQNPQ	FIKATIVHQDQAYDDKIYFFREDNPKNPEAPLN	227
Db	260	SRVAQLCRGDOGGESSLSVSKWNT	FLKAMLVCSDAATNKNFNRLQDVFLLPDPGQWRDT	319
Qy	228	SRVAQLCRGDOGGESSLSVSKWNT	FLKAMLVCSDAATNKNFNRLQDVFLLPDPGQWRDT	287
Db	320	RVYGVFSPNPNYSVAVCYSLGDI	DKVFTSLKGYHSSLNPRPGKCLPDQOPIPTETFO	379
Qy	288	RVYGVFSPNPNYSVAVCYSLGDI	DKVFTSLKGYHSSLNPRPGKCLPDQOPIPTETFO	347

Db	380	VADRHPEVAORVPMGLKPLFLHSHKYKQVAVHRMQASHGETFHVLYLTTRDGTTHKV	439
QY	348	VADRHPEVAORVPMGLKPLFLHSHKYKQVAVHRMQASHGETFHVLYLTTRDGTTHKV	407
Db	440	VEPGEQSHSAFNIETQPFERRAAAIQMSLDAERRKLYVSSQWESQVPLDLCEVYGGG	499
QY	408	VEPGEQSHSAFNIETQPFERRAAAIQMSLDAERRKLYVSSQWESQVPLDLCEVYGGG	467
Db	500	CHGCLMSRDPYCGWDQGRCSISYSSERSVLQSNPAEPHKECPNPKDPKAPLQKVS LAPN	559
QY	468	CHGCLMSRDPYCGWDQGRCSISYSSERSVLQSNPAEPHKECPNPKDPKAPLQKVS LAPN	527
Db	560	SRYLSCPMESRHATYSWRHKNENVEQSCPEGHOSPCNLCILFIENLTAQOYGHYFCEAQEGS	619
QY	528	SRYLSCPMESRHATYSWRHKNENVEQSCPEGHOSPCNLCILFIENLTAQOYGHYFCEAQEGS	587
Db	620	YFREAOHWQLLPEDGIMAE 638	
QY	588	YFREAOHWQLLPEDGIMAE 606	
RESULT 2			
ID	088371	PRELIMINARY; PRT; 393 AA.	
AC	088371;		
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DE	SEMAPHORIN L (FRAGMENT).		
GN	SEMAL.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 98389619.		
RA	LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;		
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA		
RT	viruses";		
RL	GENOMICS 51:340-350(1998).		
DR	EMBL; AF030699; G3523117; -.		
FT	NON_TER 393 393		
SQ	SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;		
Query Match 51.9%; Score 2165; DB 11; Length 393;			
Best Local Similarity 92.4%; Pred. No. 0.00e+00; Mismatches 12; Indels 1; Gaps 1;			
Db	77	HEPGSFVWVGGRGKVFHFPPECKNASVRTVNIIGSTKGCQDKQDCGNITILLERRGNG	136
QY	48	HEPGSSVWVGGRGKVFDFEPECKNASVRTVNIIGSTKGCCLKRDCENYITLLERSEG	107
Db	137	LLVCGTNARHPSCNWLNDVSVMSLGEMKGYAPSPDENSLVLPFEGDEVYSTIRKQYNG	196
QY	108	LLAGTINARHPSCNWLNVGTVP-PLGEMRGYAPSPDENSLVLPFEGDEVYSTIRKQYNG	166
Db	197	KIPFRIRGESELYTSDTVNQNPQFIKATIVHQDQAYDDKIYFFREDNPKNPEAPLN	256
QY	167	KIPFRIRGESELYTSDTVNQNPQFIKATIVHQDQAYDDKIYFFREDNPKNPEAPLN	226
Db	257	VSRVAQLCRDQGGESSLSVSKWNTFLKMLVCSDAATNRFNRLQDVFLLPDPGQWRD	316
QY	227	VSRVAQLCRDQGGESSLSVSKWNTFLKMLVCSDAATNRFNRLQDVFLLPDPGQWRD	286
Db	317	TRVYGVFSNPNWYSACVYISLGDIDRVFTSSSLKGYHMGJLNSNPRGMLCKPKQPIPTET	376
QY	287	TRVYGVFSNPNWYSACVYISLGDIDRVFTSSSLKGYHSSLPNPRPGKCLPDQQPIPTET	346
Db	377	QVADSHPEVAORVPMG 393	
QY	347	QVADRHPEVAORVPMG 363	

RESULT 3			
ID	Q64906	PRELIMINARY; PRT; 653 AA.	
AC	Q64906;		
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)		
DT	01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	SIMILAR TO GENBANK ACCESSION NUMBER L26081.		
OS	ALCELAAPHINE HERPESVIRUS 1.		
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;		
OC	GAMMAHERPESVIRINAE.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	STRAIN=C500;		
RA	ENSSER A., FLECKENSTEIN B.;		
RT	"Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";		
RL	J. GEN. VIROL. 76:1063-1067(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	STRAIN=C500;		
RA	ENSSER A., PFLANZ R., FLECKENSTEIN B.;		
RT	"Primary structure of the alcelaphine herpesvirus 1 genome.";		
RL	J. VIROL. 71:6517-6525(1997).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RX	STRAIN=C500;		
RA	ENSSER A., PFLANZ R., FLECKENSTEIN B.;		
RL	SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.		
DR	EMBL; U18243; G1000717; -.		
DR	EMBL; AF005370; G2337970; -.		
SQ	SEQUENCE 653 AA; 73645 MW; CCB53C9 CRC32;		
Query Match 42.5%; Score 1771; DB 14; Length 653;			
Best Local Similarity 46.8%; Pred. No. 0.00e+00; Mismatches 255; Conservative 95; Mismatches 184; Indels 11; Gaps 11;			
Db	102	HSLNSSDYVVGNNITVLPDFAHSSNASTALINITSTHNRHLSSTCENETILLHNQTDG	161
QY	48	HEPGSSVWVGGRGKVFDFEPECKNASVRTVNIIGSTKGCCLKRDCENYITLLERSEG	107
Db	162	LLAGTINSQKPSCW-LINNLTTOFLGPKLGLAPFSPSSGNLVLFDDQNDYTSINLYKSL	220
QY	108	LLAGTINARHPSCNWLNVGTVP-LGEMRGYAPSPDENSLVLPFEGDEVYSTIRK-QEYN	165
Db	221	GS-HKRRRTAGVELYTSOTAMHRPQVQATAVHKNESYDDKIYFFFOENSHSDFKQFPH	279
QY	166	GKIPFRIRGESELYTSDTVNQNPQFIKATIVHQDQAYDDKIYFFREDNPKNPEAPL	225
Db	280	TVPRVGQVSSDQGGESSLSVYKWTFTFLKARLACVDYDTGRIYNELQDIFIWOAPENSWE	339
QY	226	NVSRVAQLCRDQGGESSLSVSKWNTFLKMLVCSDAATNRFNRLQDVFLLPDPGQWR	285
Db	340	ETLIYGLFLSPWNFSACVVFVKDIDHVFKTSKLKNYHKLPTPRPGQCKMKNHOHVPTET	399
QY	286	DTRVYGVFSNPNWYSACVYISLGDIDRVFTSSSLKGYHSSLPNPRPGKCLPDQQPIPTET	345
Db	400	FQVADRYPEVADPYQKNAMFIIQSKYIYTKLLYRVEYG-GVFWATFYLTITKGTI	458
QY	346	FQVADRHPEVAORVPMGLKPLFLHSHKYKQVAVHRMQASHGETFH-VLYLTTRDGTI	404
Db	459	HIVRVYEDSNTTALNLEINPFQKPAPIONILLDTNLKLYVNSEWSEVSEVPLDLCVY	518
QY	405	HKVVEGEQSHSAFNIETQPFERRAAAIQMSLDAERRKLYVSSQWESQVPLDLCEVY	464
Db	519	GNDFCSFMSRDLPLCTWYNNTC-S-FK-QRVSVETGPGPANRTLSEMGCDHAPTVVYKHQV	575
QY	465	GGGCHGCLMSRDPYCGWDQGRCSISYSSERSVLQSNPAEPH-KE-CPNPKDPKAPLQKV	522
Db	576	SIPLLSNSYLSCPAVSNHADFVTKDGFTEKCHVTHKKNDCILLIANSTAINGTNTHVCN	635
QY	523	SLAPNRYLYLSCPMESRHATYSWRHKNENVEQSCPEGHOSPCNLCILFIENLTAQOYGHYFCE	582

Qy	410	-PGEHSEAFNIMEIQPPRRAAAIOTSLDAERKKLYVSQMEVSOVPLDLCEVYGGC	468
Db	522	ADCLLARDPYCAMDGHSCSRFYPTCKRRSRQDVRHGNPLTQCRGFENLKYRNAAEIVQY	581
Qy	469	HGCLMSRDPYCGWDGRCISISYSE--RSVLQISNPAEPHKECP--NRPDKAPLQKVS	524
Db	582	GVRRNSTFLECAPKSPQASIKWLLQKQ	609
Qy	525	A-PNSRYILSCPMESRHATYSWR-HKEN	550
RESULT 6			
Id	Q13214	PRELIMINARY;	PRT; 749 AA.
AC	Q13214;		
DT	01-NOV-1996	(TREMBLREL. 01, CREATED)	
DT	01-NOV-1996	(TREMBLREL. 01, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)	
OS	SEMAPHORIN V.		
OS	HOMO SAPIENS (HUMAN).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;		
OC	CATARRHINI; HOMINIDAE; HOMO.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 96210603.		
RA	SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,		
RA	ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;		
RT	"Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung		
RT	cancer deletion region and demonstrate distinct expression		
RT	patterns.";		
RL	PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).		
DR	EMBL: U28369; G974284; -		
DR	PFAM: PF00047; ig: 1.		
SQ	SEQUENCE 749 AA; 83121 MW; 6138C5D5 CRC32;		
Query Match 17.2%; Score 716; DB 4; Length 749;			
Best Local Similarity 33.0%; Pred. NO. 4.25e-148;			
Matches 167; Conservative 115; Mismatches 170; Indels 54; Gaps 41			
Db	112	ECNNEVKLLHAYNRTHLACGTGAFHPTCAFEVGHRAEPPVLRLDPGRTEGCKGKSPYD	171
Qy	93	CCENYITLLER-RSEGLLACGTNARHPSC-NNLV-N-G--TVVPL--G--EM-RGVAPES	141
Db	172	PRHRAASVLVGEELYSGVA-ADLMGRDFTIFRSIQRPSLRTEPHDSRWLNPEKFKVFW	230
Qy	142	PDNSLVLEGEDEVYTIKQYENGK-IPRRIRIGSELYT-S-DTV-MQNPQFIKAT-	196
Db	231	IPSENPDDKYYIFFRETAVEAAPALGRLSVRVGOICRNDVGGRSL-VNKWTTFLKA	289
Qy	197	IVHQDAQYDKIYIFFREDNPDKNPE-APLNVSRVAQLCRDQGGESSLSVSKWNTFLKA	255
Db	290	RLVCSVPVGEOTHDQLODFLLS-SRDH-RTPLLXAVFSTSSSIFQGSACVYVSMNDV	347
Qy	256	MLVCS--DAATNKNFNLDQVFLFDPDGGQWRDTRVYGVFSNPMN-Y--SACVYVSLGDI	310
Db	348	RRAFLGPFPAHKGPMHQWYSQGRVYPRPCMCPSKTFGTFSSTKDFPDDVTFQARNHPL	407
Qy	311	DK--V--F--RTSSL-K-GYHSSLNPNRPCKCLPQD-QPIP-TETF--QVAD--R-HPE	354
Db	408	MYNSVLPTGG-R-PLFQVQYANTYTTQIAADRVAAADGH-YDVLFTIGTDVGVTKVYISVP	464
Qy	355	VAQRVEMPGLKTPLF-H--SKYHQKVAVHRMQASHGETFHVLYLTDRGTHIKVVE-P	410
Db	465	KGSRPSAEGLLLELHVFDSDSANVTSQWISKRHLQYVASSRAVAQIALHRCAAHGRVCT	524
Qy	411	-GGEHSEAFNIMEIQPPRRAAAIOTSLDAERKKLYVSQMEVSOVPLDLCEVYGGCH	469
Db	525	ECCCLARDPYCAMDGVACTRFOPSAKRRFRRODVNRGDPSTLCSDSSRPALLEHKVFGVE	584
Qy	470	GCUMSRDPYCGWDGRCISISYSESVL--OSINPAEPHKECPNRPDKAPL-QKVS LAP	526
Db	585	GSSAFLECEPRSLQARVEMVTFQAGV	610

Qy 527 NSRYLSCPMESRHATYSWR--HKNV 551

```

RESULT 7
ID Q93018 PRELIMINARY; PRT; 750 AA.
AC Q93018
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA DANTE M.; WAMBLEY P.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL EMBL; U73167; G2880035; -.
DR PFAM; PF00047; ig; 1.
SQ SEQUENCE 750 AA; 83034 MW; CBC87E34 CRC32;

Query Match 17.2%; Score 716; DB 4; Length 750;
Best Local Similarity 33.0%; Pred. No. 4.29e-148;
Matches 167; Conservative 115; Mismatches 170; Indels 54; Gaps 41.

Db 113 ECMNFVKLLHAYNTHLLACGTGAFHGTCAFEVGHRAEPEVLRLDPRLEDGKGKSPYD 172
Qy 93 DCENYITLLER-RSEGLLACGTNARHPSC-WNLN-Y-G--TVVPL--G--EM-RGYAPFS 141
Db 173 PRHRAASVLYGEELYSGVA--ADLMGRDFTIFRSLGQRPRLTPEPHDSRWLNPKFVKVFW 231
Qy 142 PDENSLVLFGEDEVYSIRKQYNGK-IPFRRIRGESELYT-S-DIV-WQNQPIKAT- 196
Db 232 IPESENPDDKIYFFRETAVEAAPALGRUSVRVGQICRNDVGGQRSL-VNKNWTFELKA 290
Qy 197 IVHQDAQYDDKIYFFREDNPKNPE-APLNVSRVAQLCRGDQGGESSLSVSKWNTFLKA 255
Db 291 RLVCVPGVGDTHFDQLODVFLLS-SRDH-RTPLLYAVFSTSSIFQGSVAVCVYSNDV 348
Qy 256 MLVCS--DAATNKNFNLDQVFLLPDPSGQWRDTRVYGVFSNPMN-Y--SAVCVYSLGDI 310
Db 349 RRAFLGPFAPKHEGPMHQQWVSQGRVPYPRGCMCPKSTGTFSTKDFPDVIOFARNHPL 408
Qy 311 DK--V--F--RTSSL-K--GYHSLPNRPGKCLPDQ-QPIP-TEIF--QVAD--R-HPE 354
Db 409 MYSVLPTGG-R-PLFLQVGANTFTTQIADRAADRAAADGH-YDVLFTGTDYGTVLKVISVP 465
Qy 355 VAQRVEPMGPLKTLPLF-H--SKYHYQKVAVHRMQASHGETFHVLYLTLDGRGTHKVVPE-P 410
Db 466 KGSRPSAEGLLLELHVFDSEAAVTSNQISKKHQLYVASRVAQIALHRCRAHGRVCT 525
Qy 411 -GEOHSAFANIMEIQPFRRAAATQTMSLDAERKLYVSQWEVSQVPLDCEVYGGCH 469
Db 526 ECCLARDPYCAWDGVACTRFQPSAKRRFRQDVRNGDPSTLCGDSRRPALLEHKVFGVE 585
Qy 470 GCLMSPYICGWDGRCISYISDSRSLV--QSINPAEPHKECPNRPDKAPL-QKVS LAP 526
Db 586 GSSAFLECEPRSLQARVETFRAGV 611
Qy 527 NSRYLSCPMESRHATYSWR--HKNV 551

RESULT 8
ID Q99985 PRELIMINARY; PRT; 751 AA.
AC Q99985

```



```

RX MEDLINE: 94006554.
RA LUO Y., RAIBLE D., RAPER J.A.;
RT "Collapse: a protein in brain that induces the collapse and
RL paralysis of neuronal growth cones.";
RT CELL 75:217-227(1993)
DR ENBL; U02528: G410079; -.
DR PFAM; PF00047; 19; I.
SQ SEQUENCE 772 AA; 88867 MW; BC6CB9E8 CRC32;

Query Match      16.3%; Score 681; DB 13; Length 772;
Best Local Similarity 31.5%; Pred. No. 5,95e-139;
Matches 146; Conservative 119; Mismatches 150; Indels 48; Gaps 38;

Db 166 RGKSPYDPKLLTASLLVDGELYSGTA--ADPMGRDFAIFRTLGHHHPIRTEQHSRWLNDP 224
    || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 135 RGYAPFSPDENSLVLFEGDEVYSTIRKOENYNGK-IPFRRIERGESELYTS--DTV-MQNP 190

Db 225 RFTSAHLIPSDPEDDKIYFFRFNNAIDCEHTGKKATHARIGQICKNDFGGHRSLL-VNKW 283
    || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 191 QFIKATIVHQ-DQAYDDKIYFFREDNPDKNPEAPLNVSVAQLCRGDGOGESSLSVSKW 249

Db 284 TTFLKARLICVPQPGNIGIDHFDELQDVFLM-NSKOP-KNPIYGVFTTSSNIKFCSAVC 341
    ||||| :|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 250 NTFKLMLVGS--DA-ATNKNFNRLQDVFLLPPDSQGWDRTRVYGFSNPWN-Y--SVC 303

Db 342 MYSTDVRVRYFLPGAHRDGPNQWPYQCYPRGPCTPSKTGGFDSTKDLPDEVT 401
    || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 304 VYSLGDIDKVERTS-SLK-G-YH-$-S-LFNPRGKCLPDQ-QPI-PTETF--QV-A 349

Db 402 FAKSHPMYNPVPFPINS-RIMAKTDVDOYFTQIVDVRDAEDQG-YDVMFIGTDIGTVL 459
Qy 350 -DR-HPEVAQRVEPFMGPKTLPLFHSK--YHYQKVAVHRMQASHGETFHVLYLTDRGTIH 405

Db 460 KVISIPKETWHEELLEEMTVREPTVISAANKITKOOOLYIGSATGVSQLPHRCOV 519
    ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 406 KVVE-PGEQGHSAFNIME-IQFRRAAAIQTMSLDAERKLVYSOWEVSVQVPLDLCEV 463

Db 520 YGRACABCCILARDPYCAWDGSSCSRYPPTAKRTRRQDIRNGDPLTHCSDLQHHDNSGQ 579
    || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 464 YGGCGCHCLMSRDPCWGDOGRCISYSS-E-RSVL-QSINPAEPHECPN-KPKDRAP-- 518

Db 580 TLEEKIYGVENSSTFLCSPKSQRAIVY-WQFKQNDDHKVE 621
Qy 519 -LQ-KVSLA-PNSRYYLSC-PMESRHATYSWR-HKENVEQSCE 556

RESULT 14
ID Q13275 PRELIMINARY; PRT; 785 AA.
AC Q13275; Q13274;
DT 01-JAN-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HUMO.
RN [1]
RP SEQUENCE FROM N.A.
RA NELSON J., BIEWALD T.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RP SEQUENCE OF 394-436 FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE: 96210603.
RA SKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT cancer deletion region and demonstrate distinct expression
RT patterns.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR ENBL; AC000063; G1669373; -.
DR ENBL; U32172; G995788; -.
DR ENBL; U32171; G995786; -.

```

DR PFAM: PF00047; ig: 1.
SQ SEQUENCE 785 AA; 88381 MW; CD175765 CRC32;

Query Match 16.3%; Score 681; DB 4; Length 785;
Best Local Similarity 31.4%; Pred. No. 5.95e-139;
Matches 139; Conservative 122; Mismatches 139; Indels 43; Gaps 31;

Db 202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRLGKQTAMRTDOYNSRWLNDPSFIH 260
QY 139 PFPDENSELVFEGDEVYSIRKQYNGK-IPFERRIRGESELYTS-DTV-MQNPFQIK 194
Db 261 AELIPDSAEKDKLFFFRERSAE-APQSPAYVARIGRICLNDGGHCL-VNWKSTEL 318
QY 195 ATIVHQD-QAYDDKIYFFREDNPDKNPEAPLNSVRAQLCRGDGGSLSVSKWNTFL 253
Db 319 KARLVCSVPCGEDGTEHFDLQDVEV-QQTQDV-RNPVLYAVETTSVSGSVFRGSACVCYSM 376
QY 254 KAMLVCS---DAATNNKFNRLQDVLDPDPSGQWRDTRVIGVFSNP-WNY---SACVYSL 307
Db 377 ADIRMFVNGFPAHKEGPNYQWPFSGKMPYPRPGTCPGGTFTPSMKSTKDYDPDEVINMR 436
QY 308 GDIDKVFRTS-SLK-G-YH-----SS-LPNPRGKCLPDQ-QP-I-PTETF--QVAD--R 351
Db 437 SHPLMQAVYPLQRRPLVVRT-GAPYRLTTIAVDQVDAADG-RYEVLFGLTDRGTGVOKVI 494
QY 352 -HPEVAQRVEPMG--PLKTPLFHSHYKQVAVHRMQASHGETFHVLYLTTRDGTIRKVV 408
Db 495 VLPKDDQLEELLEEVEYFKDPAPKVTYTISSKROOLYVASAVGYTHLSLHRCQAYGAA 554
QY 409 EPGEQEHSFA-FNIMEIQPFRRAAQTMSLDRERKLVSOWEVSQVPLDCEVYGGG 467
Db 555 CADCCCLARDPYCAWDGQACSRYSATSKRRSRRODVRHGNPIROCGRFNSNANKNAVESVQ 614
QY 468 CHGCLNSRDPCYCGWDGRCIS-IYSSER-SVLOSINPAEPHKCP--NPKPKAPLOKYS 523
Db 615 YGVAGSAFLFECQPRSPQATVKW 637
QY 524 LA-PNSRYLLSCPMSRHSATYSW 545

RESULT 15
ID Q15704 PRELIMINARY; PRT; 785 AA.
AC Q15704;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 96226360.
RA ROCHE J., BOLDGF., ROBINSON M., ROBINSON L., VARELLA-GARCIA L.,
RA SWANTON M., WAGGONER B., FISHEL R., FRANKLIN W., GEMMILL R.,
RA DRABIN H.
RT "Distinct 3p21.3 deletions in lung cancer and identification of a new
RT human semaphorin."
RL ONCOGENE 12:1289-1297(1996).
RL EMBL; U33920; G1000207; -.
DR PFAM: PF00047; ig: 1.
SQ SEQUENCE 785 AA; 88385 MW; 9971F14E CRC32;

Query Match 16.3%; Score 679; DB 4; Length 785;
Best Local Similarity 31.6%; Pred. No. 1.98e-138;
Matches 140; Conservative 121; Mismatches 139; Indels 43; Gaps 31;

Db 202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRLGKQTAMRTDOYNSRWLNDPSFIH 260
QY 139 PFPDENSELVFEGDEVYSIRKQYNGK-IPFERRIRGESELYTS-DTV-MQNPFQIK 194
Db 261 AELIPDSAEKDKLFFFRERSAE-APQSPAYVARIGRICLNDGGHCL-VNWKSTEL 318

Qy 195 ATIVHQD-QAYDDKIYFFREDNDKKNPEAPLNYSRVAQLCRGQGGESSLSVSKWNTFL 253
Db 319 KARLVCSVPGEDGTETHFDELQDVV-QQODV-RNPVIYAVETSSGVSFRGSVAVCYSM 376
Qy 254 KAMLVCS--DAATNKNFNRLQDVFLLPDPGQWRDTRVYGVFSNP-WNY--SAVCVYSL 307
Db 377 ADIRMFNGPFAHKEGPNYQWMPFSGKMPYRPGTCPCGGTFTPSMKSTKDYPDEVINFM 436
Qy 308 GDIDKVFRTS-SLK-G--YH---SS-LPNRPGCKLDPQ-OP-I-PTETF--QVAD--R 351
Db 437 SHPLMYQAVYPLQRRPLVVRT-GAPYRLTTIAVDQVDAGDG-RYEVLFLGTDRGTOKVI 494
Qy 352 -HPEVAQRVEPMG--PLKTPLFHSHKYHYKVAVHRMQASHGETPHVLYLTDRGTIHKV 408
Db 495 VLPKDDQEMEELMLEEVEVFKDPAPVKMTITSSKRQOLYVASAVGVTHLSIHRCOAYCAA 554
Qy 409 E-PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERRKLYVSSQWEVSOVPLDLCEVYGG 467
Db 555 CADCCCLARDPYCAWDGQACSRYTASSKRRSRQDVHRHGNPIROCGRFNSNANKNAVESVQ 614
Qy 468 CHGCLMSRDPYCGNDQGRGIS-IYSSER-SVLQSIINPAEPHKECP--NPKPDRAPLQKVS 523
Db 615 YGVAGSAFLECPQPRSPQATVKW 637
Qy 524 LA-PNSRYLLSCPMESRHHATYSW 545

Search completed: Thu Jul 8 19:27:24 1999
Job time : 45 secs.

WQELH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:32:13 1999; MasPar time 25.45 Seconds
502.244 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
(1-601) from US09041236.pap (36 of 45)

Perfect Score: 4488
Sequence: 1 LLLLLWAAASAGHLRSGP.....EABGGSYFREAQHWQLLPED 601

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266508 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.619; Variance 154.932; scale 0.236

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	661	14.7	771	R71380	Human semaphorin III	8.11e-51
2	620	13.8	477	R71375	Human semaphorin	8.06e-47
3	612	13.6	775	W63748	Human semaphorin	4.84e-46
4	606	13.5	861	W17658	Mouse CD100 antigen	1.85e-45
5	599	13.3	861	W58540	Human semaphorin	1.88e-45
6	544	12.1	776	W51313	Rat semaphorin W	1.89e-39
7	544	12.1	862	W17657	Human CD100 antigen	1.89e-39
8	463	10.3	411	R71381	Vaccinia virus semaph	1.15e-31
9	459	10.2	780	R71379	Grasshopper semaphori	2.76e-31
10	428	9.5	537	W51314	Human semaphorin W	2.47e-28
11	390	8.7	974	W64221	Human secreted protei	9.73e-25
12	378	8.4	712	R71384	Tribolium semaphorin	1.31e-23
13	367	8.2	650	R71382	Drosophila semaphorin	1.41e-22
14	362	8.1	724	R71383	Drosophila semaphorin	4.15e-22
15	352	7.8	930	W57260	Human semaphorin Y	3.57e-21
16	325	7.2	888	W19857	Human semaphorin Z	1.15e-18

17	316	7.0	887	25	W19856	Rat semaphorin Z.	7.83e-18
18	296	6.6	929	32	W57259	Rat semaphorin Y.	5.40e-16
19	157	3.5	122	13	R71385	Varola major virus s	8.71e-04
20	107	2.4	111	32	W51315	Human semaphorin W pa	7.05e-00
21	101	2.3	2329	25	W25038	Partial BRCA2 cancer	1.91e-01
22	99	2.2	591	4	R23006	Protein transcribed f	2.66e-01
23	98	2.2	832	33	W61092	Taq DNA polymerase I	3.13e-01
24	96	2.1	832	33	W61090	Taq DNA polymerase I	4.32e-01
25	93	2.1	4572	30	W52845	A. mediterranei rifam	6.99e-01
26	91	2.0	478	1	R04881	Recombinant elastase.	9.58e-01
27	89	2.0	534	20	W25031	Partial BRCA2 cancer	1.31e-02
28	89	2.0	554	13	R66209	Novel thermostable DN	1.31e-02
29	91	2.0	598	9	R48631	Sequence of nuclear r	9.58e-01
30	90	2.0	638	5	R03924	E. coli HSP (dnaK).	1.12e-02
31	89	2.0	680	4	R23143	Mutant thermostable D	1.31e-02
32	92	2.0	832	33	W61091	Taq DNA polymerase I	8.18e-01
33	89	2.0	832	13	R76693	DNA-polymerase (F73L,	1.31e-02
34	89	2.0	832	1	P90556	Purified native therm	1.31e-02
35	89	2.0	832	33	W61087	Taq DNA polymerase I	1.31e-02
36	89	2.0	832	33	W61088	Taq DNA polymerase I	1.31e-02
37	89	2.0	832	13	R76690	Taq DNA-polymerase RE	1.31e-02
38	89	2.0	832	33	W61089	Taq DNA polymerase I	1.31e-02
39	89	2.0	833	36	W59942	Amino acid sequence o	1.31e-02
40	89	2.0	833	27	W24211	Cleavage DN nuclease.	1.31e-02
41	89	2.0	833	27	W24212	Cleavage DA nuclease.	1.31e-02
42	92	2.0	854	30	W56309	Class II S-receptor ki	8.18e-01
43	92	2.0	855	6	R29815	S receptor kinase pro	8.18e-01
44	89	2.0	3418	26	W19211	Human breast cancer s	1.31e-02
45	89	2.0	3418	26	W23287	Human breast and ovar	1.31e-02

ALIGNMENTS

RESULT 1
ID R71380 standard; Protein; 771 AA.
AC R71380;
DT 21-NOV-1995 (first entry)
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN WO9507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994: U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB: Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 60-63; 101pp; English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II,
CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA:

Query Match 14.7%; Score 661; DB 13; Length 771;
Best Local Similarity 30.3%; Pred. No. 8.11e-51;
Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;

DB 166 rgkspdyklltasllidgelysgta-adfmgdrfafrtlghhphirteqhsrwlndp 224

QY 93 DCENYITLLER-RSEGLACGYNARHPSG--WNLVNGTVVPLGEN-RGYAPSPDENSLV 148
Db 167 vnnvgelysgts-yfnglsepilistnssnshpirtayaipwlnepsvfadvaiqlkspdgpe 225
QY 149 LFEQDEVYSTIRKQYNGKIPRFRIRISESELYTSDTV--MONPOFIKATIVHQ--D-Q 202
Db 226 geddkvyyfftevsveyefvflmiprvarcvkgdggglrtliq-kkwtstflkarlicskp 284
QY 203 AYDDKIYFFREDNDPNKPEAPLNVSRAQLCRGQGGESSLSVSKWNTFLKAMLVCSDA 262
Db 285 dsglvfnllqdvvlrap-gl-kepvfayvfpqlnngvlsavceaytlatveavfsrky 342
QY 263 ATNKNENRLQDVFLLPDPGSGWRDTRVYGVFSNPWY---SACVYSLGDDIKVF-R--- 315
Db 343 mgsatveqshktwryngvpvptprpqacldsearaanytsslnlpdktlqfvkdhplmd 402
QY 316 --TSSL-----K-G-YHSSLNPRGKCL-PDQQIP-TETFOVADRHPE-V-AOR-VE- 360
Db 403 svtpidnprklikdvnytiqivdrtdqldgtfydvdmfistdrgalhkaviltkev--v 460
QY 361 PMGPLKT-P-LPHSKYHYQKVAHVMAQASHGTFHVLVLTDRGTIHKVVEPGEQHSFA 418
Db 461 --iectqlfrdfepvltlllskkrkrfvyagsnsvvgaplafacekhgs-cedcvlard 517
QY 419 FNIMEIQPFRRAAAIQTMSLDAER-RKL-YVSSQWEVSVQVPLDLCEVYGGCHGCLMSRD 476
Db 518 pycawspalkacvtlhqeeas 538
QY 477 PYCGWDQG-R-CISYSSERS 495

RESULT 6

ID W51313 standard; Protein; 776 AA.
AC W51313;
DT 08-SEP-1998 (first entry)
DE Rat semaphorin W.
KW Rat; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW immunosuppressant; gene therapy; diagnosis; research reagent.
OS Rattus norvegicus.
PN W09815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PA 09-OCT-1996; JP-287636.
PR (SUNU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-261015/23
DR N-PSDB: V07279, V07280.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Claim 1: Page 60-64; 90pp; Japanese.
CC The present sequence represents rat semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 776 AA;

Query Match 12.1%; Score 544; DB 32; Length 776;
Best Local Similarity 27.4%; Pred. No. 1.89e-39;
Matches 163; Conservative 145; Mismatches 232; Indels 55; Gaps 39;

Db 26 lllllailsapcvgrsvrptsipiseadsyltrfaashtynysallvdpashtlyvg 85
QY 1 LLLLLLWAAAASQAQHL-RSGPRI-FAVKNKGHVQDRVDFGQTPEHTVLFHEPGSSVWVG 58
Db 86 ardsifaltlpsgerpriidwmpv-ethrqncrkkgkedecnfldilaivnashllt 144
QY 59 GRGKYV-L-FDFPEGKNASV-RTVNIGSTKGSCL--DKR-D-CENYITLLERSEGL-LLA 110

Db 145 cgtfafdpkcgvidvssfqvveriesgrgkcpfopagrsaavmaggvlytatvk-nflgt 203
QY 111 CGTNARHPSCNNL-VNG--TVVPLGEMRGYAFSPDENSLVLFEGDEVYSTIRKQYNGK 167
Db 204 epilisravgraedwirutetlsswlnapafvaamvlspeewgdgddeiffettetsvl 263
QY 168 IPRFRIRIGESELYT-SDTV---MONPOFIKATIV-H-Q--DQAYDDKIYFFREDNDPK 219
Db 264 dsyerikvprvarcagdlgrktliq-grwtftikadilcpqgehrasgvlqamaelrp 322
QY 220 NPEAPLNVSRAQLCRGQGGESSLSVSKWNTFLKAMLVCSDAATNKNENRLQDVFLLPD 279
Db 323 qpqa-gtpifgyfissqwegaaaisavcafrpqdiravngpfrelkhdcnrglpvmndev 381
QY 280 PSGQWRDTRVYGVFSNPWN---YSACVYSLGDDIKVFRTS--SLK-GYHSSLP--N-- 328
Db 382 pqprrpgeciannmkllqfgsslsipdrvlftirdhplmdrvpfpadgrpllvtdtaylor 441
QY 329 --PRPGKCLPDQQIP--TETFOVADRHPE-V-AOR-VE-PMGPLKT-PLFHSK-YHYQK 378
Db 442 vvahrvtslsqkeydvlylgtedghllhravrigaq-ls-vledlal--fpepqvvesmkl 497
QY 379 VAVHMAQASHGTFHVLVLTDRGTIHKVVEPGEQHSFAFNIMEIQPFRRAAAIQTMSL 438
Db 498 yhdw--llvgshtvqntsnscrlqs-csecilagdpvcawsfridacvahagehrm 554
QY 439 DAERRKLYVSSQWEVSVQVPLDLCEVYGGCHGCLMSRDYCGWD-Q-GRCISYSSERSV 496
Db 555 vqdliesadvslcpkegehpvvfevpvavtvghvlpccspassawascvwhapsv 609
QY 497 LQSNAPEPHKECPNPKPDKAPLOKVLAPNSRYLLSCPMESRHATYSWRHKENY 551

RESULT 7

ID W17657 standard; Protein; 862 AA.
AC W17657;
DT 24-JUL-1997 (first entry)
DE Human CD100 antigen.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
KW vaccine.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..41
FT /label= sig_peptide
FT protein 42..862
FT /label= Mat_protein
FT domain 42..553
FT /label= Semaphorin_domain
FT domain 554..630
FT /label= Ig-like_domain
FT domain 631..733
FT /label= Stalk_domain
FT domain 735..752
FT /label= Transmembrane_domain
FT domain 753..862
FT /label= Cytoplasmic_domain
FT modified_site 808..815
FT /label= Phosphorylation
FT /note= "putative tyrosine phosphorylation site"
PN W09717368-A1.
PD 15-MAY-1997.
PF 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422
PA (DAND) DANA FARBER CANCER INST.
PI Boussioutis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB: T60665.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Claim 7; Page 70-72; 135pp; English.
CC Human CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell

[illegible]

Qy 208 IYFFREDNPKNEAPLNVSRVQLCRGQOGGES-SLSVSKWNTFLKAMLVCSDAATNK 266
Db 294 tfy-f-dvlqaltgpnvlhgrsa-lfgvfttqtntsiptsavcafyldeiergfeqkfkeq 350
Qy 267 NFNRLOQVFL-LPDPGQ-QWRDTRYGVFSNPNW-Y--SACVYSLGDIKVF--R-TS- 317
Db 351 rslgawtpvsedrypsrpsacagvgaaalfssrdlppdvtltfikahpildavppvt 410
Qy 318 -SLKG-Y-H-S-LFNPRPKCLP-DOQPIETTFQVADRHPE-V-AORV-EP-MGPLK 366
Db 411 hqpllttsralltqavdgmagphsn-itvmflgsndgtvklvtpggrsggpepille 469
Qy 367 -TPLFH--SKYHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVVEPGEQHSFAFNIME 423
Db 470 eldaysparcsgkrtaqtarrilgleldteghrlfvaosgivlplsrcarhga-cqrs 528
Qy 424 -IQPF---R---RAA--A--IQTMSLDAERRKLYVSSQWEYSQVPLDLCEVYGGCH-G 470
Db 529 clasgdpvcgwhsrgcvgdirsggtdvdaqnqesmehgdcqd 572
Qy 471 CLMSRDPYCGWDQGR-CISYSS-RSVLQSTINPAE-PHKECPN 511

Search completed: Thu Jul 8 19:33:14 1999
Job time : 61 secs.

THIS PAGE BLANK (USPTO)

WIRELESS

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:31:21 1999; MasPar time 27.35 Seconds
Tabular output not generated. 880.457 Million cell updates/sec

Title: >US-09-041-236-2
Description: (1-601) from US09041236.pep (36 of 45)
Perfect Score: 488
Sequence: 1 LLLLLWAAAASAGHLRSGP.....EAQEGSYFRAQHWWLLPED 601

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40086593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 49.338; Variance 86.803; scale 0.568

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	1883	42.0	653	2 T03102	semaphorin homolog A3	0.00e+00
2	737	16.4	748	2 I48744	semaphorin A - mouse	1.00e-134
3	725	16.2	751	2 I48748	semaphorin E - mouse	5.91e-132
4	716	16.0	749	2 G01856	semaphorin V - human	7.03e-130
5	688	15.3	753	2 G02173	semaphorin III family	1.96e-123
6	681	15.2	772	2 A90699	collapsin - chicken	7.93e-122
7	667	14.9	666	2 I58169	semaphorin III - mouse	1.29e-118
8	669	14.9	772	2 I48747	semaphorin D - mouse	4.50e-119
9	661	14.7	771	2 D49423	semaphorin III precursor	3.07e-117
10	618	13.8	782	2 I48746	semaphorin C - mouse	2.06e-107
11	522	11.6	834	2 S66498	M-sema F protein prec	1.09e-85
12	480	10.7	760	2 I48745	semaphorin B - mouse	2.64e-76
13	470	10.5	403	2 E42521	A39R protein - vaccin	4.41e-74
14	463	10.3	441	2 S29921	hypothetical protein	1.57e-72
15	461	10.3	730	2 JH0798	fasciclin IV precursor	4.36e-72
16	448	10.0	1074	2 J65928	semaphorin F precursor	3.27e-69
17	389	8.7	711	2 A49423	semaphorin I precursor	2.68e-56
18	367	8.2	656	2 D49423	semaphorin I - fruit	1.51e-51
19	362	8.1	724	2 C49423	semaphorin II precursor	1.79e-50
20	293	6.5	295	2 J01775	Sall9R protein - vacc	6.56e-36
21	196	4.4	1884	2 J64973	plexin 3 precursor -	1.21e-16
22	170	3.8	1872	2 J64976	plexin 3 precursor -	7.07e-12
23	171	3.8	1905	2 I51553	Plexin - African claw	4.68e-12

ALIGNMENTS

RESULT 1
ENTRY #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999

ACCESSIONS T03102
REFERENCE Z14840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-653 #label ENS
#cross-references EMBL:AF005370; NID:G2337967; PID:G2337970
SUMMARY #length 653 #molecular-weight 73645 #checksum 5501

Query Match 42.0%; Score 1883; DB 2; Length 653;
Best Local Similarity 46.3%; Pred. No. 0.00e+00;
Matches 274; Conservative 105; Mismatches 200; Indels 13; Gaps 13;

Db	57	LLML-SAITAAKSRFDKPLIVNLTGFGQHRF-FGQPEPHTVLFHSLNDSDDYVGVGN	114
Qy	1	LLLLLWAAAASAGHLRSGPRIFAVWKGHVGDVDFGQTEPHTVLFHEPGSSVWVGG	60
Db	115	NTIYLFDFAHSSNASTALINTSTHNTLRSLSTCENITLLHNTDGLAGTNSQRPSC	174
Qy	61	GKVLDFPEEGKASVTRVNIQSTGKCLDKRDCENITLLERSEGLAGTNAHPSD	120
Db	175	W-LNNLTTLQPLGLAPSPSSGNLVLFDDNDTSTINLYKSLGS-HKFRRIAGQV	232
Qy	121	WNLVNGIWPV-LGEMRGVAPSPDENSLVFEQDEVSTIRK-QEYNGKIPFRIRGES	178
Db	233	ELYTSDTAMHRPQVQAVAHKNSYDDKIYFFFOENSHSDFKQPHPTVPRVGVCSDDQ	292
Qy	179	ELYTSDTVMQNPQIKATIVHQDQYDDKIYFFREDNPKNPEALNVSAVQLCRDQ	238
Db	293	GGESLSVYKWTTLKARLACVDYDGRVNELODIWIWAPENSWEETLYIGFLSPWN	352
Qy	239	GGESLSVSKWNTFLKAMLVCSDAATNKNRQLQDFLLDPDPSGQWRDTRYGVFSNPN	298
Db	353	FSAYCVFTVKDIDHVFETSKLKNVHKLPTRPGCKNKHQHVPTETFOVADRYPEVADP	412
Qy	299	YSAYCVVSLGDIKVFRTSSLUKGVHSSLPNRPGRKLPDQDQPIPTETFOVADRYPEV	358
Db	413	VYQKNAMFPIQSKYIVTKLLVYRVEYG-GVFWATIFYLTIKGTTHIYVRYEDSNSTT	471

QY 304 VYSLGDIKVERTS-SLK-G--YH--S--S-LPNPRPKCLPDO-QPI-PTETF--QV-A 349

Db 402 FARSHPAMYNVPFINS-RPIMIKTVDYQFTQIVVDRVDAEDGO-YDVWFIGTDIGTIVL 459

QY 350 -DR-HPEVAQRVEMGPLKPLFHSK--YHYQKVAVHRMQASHGETHVLVLTDRGTIH 405

Db 460 KWSIPKETWHELEEVLEEMTVREPTVISAMKISTKQOOLYIGSATGYSQVPLHRCDV 519

QY 406 KVVE-PGEQHSFAFNIME-IQPFRAAAIQTMSLDAERKRLYVSSQWVPLDLCEV 463

Db 520 YGKACAECCCLARDPYCAWDGSSCSRYPFTAKRTRRQDIRNGDPLTHCSDLQHHNDPSGQ 579

QY 464 YGGCHGCLMSRDPYCGWDGRCISYSSSE-RSVL-QSINPAEPHKECPN-PRPKDAP-- 518

Db 580 TLEKIITYGVENSSTFLECSPKSORAIVY-WQFQKNDHKEV 621

QY 519 -LO-KVSLA-PNSRYYLSC-PMESRHATYSWR-HKENVEQSC 556

RESULT 7

ENTRY semaphorin III - mouse (fragment)

TITLE #formal_name Mus musculus #common_name house mouse

ORGANISM 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change

DATE 04-Sep-1998

ACCESSIONS I58169

REFERENCE Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;

#authors Tessler-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.

#journal Neuron (1995) 14:949-959

#title Semaphorin III can function as a selective chemorepellent to

#cross-references MUID:95267432

#accession I58169

##status preliminary; translated from GB/EMBL/DBJ

##molecule_type mRNA

##residues 1-666 #label RES

##cross-references GB:L40484; NID:g703189; PID:g703190

GENETICS SemIII

CLASSIFICATION #superfamily semaphorin

SUMMARY #length 666 #checksum 9654

Query Match 14.9%; Score 667; DB 2; Length 666;

Best Local Similarity 30.1%; Pred. No. 1.29e-118;

Matches 150; Conservative 130; Mismatches 168; Indels 50; Gaps 39;

Db 60 RGKSPDPKLLTASLLDGLYSGTA-ANEMGRDFAIFRTLGHHPHIRTQHDNRMLNDP 118

QY 135 RGVAFFSPDENSLLVFEQDEVYSTIRKQYNGK-IPFRIRIRGESELYTS--DTV-MQNP 190

Db 119 RFISAHILIPESDNPDDKVVFFRENAIDGESHGKATHARIGQICKNDFGGHSL-VNKKW 177

QY 191 OFIKATLVHO-DOAYDDKIYFFREDNPKNEAPLNVSRVAQLCRDQGGESSLSVSKW 249

Db 178 TTFKLARLICSVPNGIDTHFDELOQVFLM-NSKOP-KNPIYGVFTTSSNIFKGSVAVC 235

QY 250 NTFKLAMLVCS--DA-ATNNKFNRLQDVFLPDPSCQWRDTRYGVFSNPWN-Y--SAVC 303

Db 236 MYSMSDVRVLLGPGYAHRDGPNYQWVPGYVPRPGTCSKTFGGFDSKLPDDVIT 295

QY 304 VYSLGDIKDV---F--RTSS-LK--CYHSSLNPNRPGKCLPDO-QPI-PTETF--QV-A 349

Db 296 FARSHPAMYNVPFINN-RPIMIKTDVNYQFTQIVVDRVDAEDGO-YDVWFIGTDVGTIVL 353

QY 350 -DR-HPEVAQRVEMGPLKPLFHSK--KYHYQKVAVHRMQASHGETHVLVLTDRGTIH 405

Db 354 KVSVPKETWHDLEEVLEEMTVFREPTTISAMELSTKQOOLYIGTAGVAQLPHRCIDI 413

QY 406 KVVE-PGEQHSFAFNIME-IQPFRAAAIQTMSLDAERKRLYVSSQWVPLDLCEV 463

Db 414 YGKACAECCCLARDPYCAWDGSSCSRYPFTAKRTRRQDIRNGDPLTHCSDLQHHNDHGP 473

QY 304 VYSLGDIKDV---F--RTSS-LK--CYHSSLNPNRPGKCLPDO-QPI-PTETF--QV-A 349

Db 412 PYFPINN-RPIMIKTDVNYQFTQIVVDRVDAEDGO-YDVWFIGTDVGTIVLVVSPKETW 469

QY 358 RVEPMGLTPLFHS--KYHYQKVAVHRMQASHGETHVLVLTDRGTIHKVVE-PGEGE 414

Db 470 HDLEEVLEEMTVFREPTTISAMELSTKQOOLYIGTAGVAQLPHRCIDIYKACAECCL 529

QY 415 HSFAFNIME-IQPFRAAAIQTMSLDAERKRLYVSSQWVPLDLCEVYGGCHGCLM 473

Db 530 ARDPYCAWDGSSCSRYPFTAKRTRRQDIRNGDPLTHCSDLQHHNDHGPSLEERTIYGV 589

QY 474 SRDPYCGWDGRCISYSSSE-RSVL-QSINPAEPHKECPN-PRPKD--AP-LQ-KVSLA- 525

QY 464 YGGCHGCLMSRDPYCGWDGRCISYSSSE-RSVL-QSINPAEPHKECPN-PRPKD--AP 518

Db 474 SLERIIYGVENSSTFLECSPKSORALVWQFORNRDKEEIKMGDHIIRTEQGLLRS 533

QY 519 -LO-KVSLA-PNSRYYLSC-PMESRHATY-SW-RHKENVEQSCPEG-HQ-SPNCILFIEN 570

Db 534 LQKKDSGNLYLCHAVEHGF 551

QY 571 LTAQQYGHYFCEAQEGSY 588

RESULT 8

ENTRY semaphorin D - mouse

TITLE #formal_name Mus musculus #common_name house mouse

ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change

DATE 04-Sep-1998

ACCESSIONS I48747

REFERENCE Puschel, A.W.; Adams, R.H.; Betz, H.

#authors Neuron (1995) 14:941-948

#journal Marine semaphorin D/collapsin is a member of a diverse gene

#title family and creates domains inhibitory for axonal extension.

#cross-references MUID:95267431

#accession I48747

##status preliminary; translated from GB/EMBL/DBJ

##molecule_type mRNA

##residues 1-772 #label RES

##cross-references EMBL:X85993; NID:g854329; PID:g854330

GENETICS semD

CLASSIFICATION #superfamily semaphorin

SUMMARY #length 772 #molecular-weight 88710 #checksum 1776

Query Match 14.9%; Score 669; DB 2; Length 772;

Best Local Similarity 30.2%; Pred. No. 4.50e-119;

Matches 170; Conservative 143; Mismatches 182; Indels 67; Gaps 51;

Db 58 HTFLDE-ERSRLYVCAKDHIFSFLVNLVNIKDFQKIWPVSY-TRRDECKWAGKDILKECA 115

QY 43 HTVLFHEPGSSVWVGGRKVLDFDPBGN-AS-VRTNIGSTKGC--LQK---RDCE 95

Db 116 NFKVLEAYNOTHLXACGTGAFHPICITYIEVGHHPEDNIFKLQDSHFENGKSPDPKLL 175

QY 96 NYITLLERSEG-LLACGTNARHPSCWNL-V--N--GTVVPL--G--EM-RGVAFSPDE 144

Db 176 LTRASLIDGELYSGTA-ADFMGRDFAIFRTLGDHHPDIPTEQHDNRMLNDRPFIASHLPE 234

QY 145 NSLVLFEGDEVYSTIRKQYNGK-IPFRIRIRGESELYTS--DTV-MQNPQFIKATIVHQ 200

Db 235 SDNPEDDKVYFFRENAIGGESHGKATHARIGQICKNDFGGHSL-VNKKWTFKLARLIC 293

QY 201 -DOAYDDKIYFFREDNPKNEAPLNVSRVAQLCRDQGGESSLSVSKWTFKLAMLYC 259

Db 294 SVYGPNGIDTHFDELOQVFLM-NSKOP-KNPIYGVFTTSSNIFKGSVAVCYMSMSDVRV 351

QY 260 S--DA-ATNNKFNRLQDVFLPDPSCQWRDTRYGVFSNPWN-Y--SAVCVYSLGDIKV 313

Db 352 FLGPAHRDGPYQWVPGYVPRPGTCSKTFGGFDSKLPDDVITFGRSHPAMYN 411

QY 314 FRTS-SLK-G--YH--S--S-LPNRPGKCLPDO-QPI-PTETF--QV-A-DR-HPEVAQ 357

Db 412 PYFPINN-RPIMIKTDVNYQFTQIVVDRVDAEDGO-YDVWFIGTDVGTIVLVVSPKETW 469

QY 358 RVEPMGLTPLFHS--KYHYQKVAVHRMQASHGETHVLVLTDRGTIHKVVE-PGEGE 414

Db 470 HDLEEVLEEMTVFREPTTISAMELSTKQOOLYIGTAGVAQLPHRCIDIYKACAECCL 529

QY 415 HSFAFNIME-IQPFRAAAIQTMSLDAERKRLYVSSQWVPLDLCEVYGGCHGCLM 473

Db 530 ARDPYCAWDGSSCSRYPFTAKRTRRQDIRNGDPLTHCSDLQHHNDHGPSLEERTIYGV 589

QY 474 SRDPYCGWDGRCISYSSSE-RSVL-QSINPAEPHKECPN-PRPKD--AP-LQ-KVSLA- 525

```
Db 590 ENSSTFLECPKSRQALVY-WQ 610
QY 526 PMSRYLSC-PMESRHATYSWR 546

RESULT 9
ENTRY #type complete
TITLE semaphorin III precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
#accession D49423
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-771 #label KOL
#cross-references GB:L26081; NID:g799328; PID:g436360
GENETICS
#gene GDB:SEMA1
#cross-references GDB:283448
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 771 #molecular_weight 88889 #checksum 6249

Query Match 14.7%; Score 661; DB 2; Length 771;
Best Local Similarity 30.3%; Pred. No. 3,07e-117;
Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;

Db 166 GKSPYDPKLLTASLLDGLYSGTA-ADFMGRDPAIFRTGLHHHPITEQHSRWLNDP 224
QY 135 RGYAPSPDENSESLVFEDEVYSTIRKQYNGK-IPFRFRIRGESELYTS--DTV-MQNP 190
Db 225 KFISAHLLSESDNPEDDDYFFERENADIGHSKATARTGICQKNDFFGGHRS-L-VNKK 283
QY 191 QFIKATIVHQ-DOAYDDKIIYFFREDNPKNPEAPLNSRVAQLCRGQGGESSLSVSKW 249
Db 284 TTFKLARLCSVPGNGIDTHFDELQDVFLM-NEKDP-KNPVYGVFTTSSNIFKGSVAC 341
QY 250 NTFKLAMLVS--DA-ATNKNENRLQDVFLLPDPSSQWRDTRVYGFSPNWN-Y--SVC 303
Db 342 MYMSDVRVFLGPAHRDGNQVWPYQGRVYPRPGTCPSKTEGGDSTKDLPPDDVIT 401
QY 304 VYSLGDIDKVFRTS-SLK-G-YH--S-S-LPNRPKGKCLPDQ-OPI-PTETF--QV-A 349
Db 402 FARSHPMYNPVFPNNRPVITKTDVNYQFTQIVVDYDAEDGQ-YDYNFICTDVGTVLK 460
QY 350 -DR-HPEVAQRVEPMGPKTLP-FHSKYHYQKVAHRMQASHGETFHVLYLTDRGTIHK 406
Db 461 VYVIPKETWYDLEEVLLBEEMTVFREPTAISAMELSTKQQLYIGSTAGVAOLPLHRCDIY 520
QY 407 VVE-PGEQHSFAFIME-IQPFRAAAIQTMSLDAERKLYVSSQWESQVPLDCEVY 464
Db 521 GKACAECLARDPYCAMDGSACSRVEPTAKRTRRQD-IRNGDPLTHCSDLHNDHHGSP 580
QY 465 GGGCHGLMSRDYPCGWQDGRCSISYSE-RSVL-QSINPAEPHKECPNPKPKD---AP 518
Db 581 EERIIVGVENSTFLECPKSRQALVYVQFORNEERKEETRVDDHIIRTDGGLLLRSQ 640
QY 519 LOKVSLA-PNSRYLYLSC-PMESRHATY-SW-RHKENV-EQSCPEGHQ-SPNCILFIENLT 572
Db 641 OKDSGNYLCHAVEHGFIQ 658
QY 573 AQQYGHYFCERQEGSYFR 590

RESULT 10
ENTRY #type fragment
TITLE semaphorin C - mouse (fragment)
```

```
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
#accession I48746
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-782 #label RES
#cross-references EMBL:X85992; NID:g854327; PID:g854328
GENETICS
#gene semC
SUMMARY #length 782 #checksum 1571

Query Match 13.8%; Score 618; DB 2; Length 782;
Best Local Similarity 33.7%; Pred. No. 2,06e-107;
Matches 149; Conservative 96; Mismatches 148; Indels 49; Gaps 32;

Db 74 KRDCQNYIKILLPLNSSHLLTCGTAAPSPLCAYTHIASFTLAQDEAGNVILEDKGHCPF 133
QY 91 KRDCENYIT-LLERRSEGLLACGNARHPSC-W-N-----LV-NGTV-VPLGEMHGYAPF 140
Db 134 DPNFKSTALVVDGELYTGVS-SFGNDPAISRQSSRPTKTESLNLWLODPAFVASATS 192
QY 141 SPDENSLVFEDEVYSTIRKQYNGKIPFRFRIRGESELYTSDTV--MQNPQFIKATIV 198
Db 193 PESLGSPTGDDDKYFFSETGQEFFEFTNI-VSRVARVCKDEGGERVLO-QRWTSFL 250
QY 199 HQDQ-A-Y--DDKIYFFREDNPD-KNPEAPLNSRVAQLCRGQGGESSLSVSKWTF 253
Db 251 KAOLLCSRPDDGEPNVLDVFTLNPNPQ-DWRKTLSTGVTESQWHRGTGSAICVFTM 309
QY 254 KAMLVCSDAATNKNFNRLQDVFL-L-PDPSQWRDTRVYGFSPNWN---Y--SVCVYSL 307
Db 310 NDVOKAFDGLYKYNRETQWYTHQVTPRPGACIATNSARERKINSLSLOLPDRLVNL 369
QY 308 GDIDKVF-----RTSS-LKG-Y-HSS-LPNRPKGKCLPD--QQPIPTTFQVADRHPEVA 356
Db 370 KDHFMDQVRSRLLLOPRARYQVAVHRVPGHLS-TYDVLFLGTGDRLHKAVTLSSR 428
QY 357 QRVEPM-GPLKTPLF--HSKYHYQKVAHRMQASHGETFHVLYLTDRGTIHKVVEPGEQ 413
Db 429 VH--I-IEELQIFPOGQVQNLDDLHSHGLLYASSHSGVVQVVPVANCPLYPT-CGDCLL 483
QY 414 EHSFAFIMEIQPFRAAAIQTMSLDAERKLYVSSQWESQVPLDCEVYGGGCHGLM 473
Db 484 ARDPYCAWTGSACRLASLYQPD 505
QY 474 SRDPYCGWDQGRG-I-SIYSSE 493

RESULT 11
ENTRY #type complete
TITLE M-sema F protein precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change
#accession S66498
#status preliminary
#molecule_type mRNA
#residues 1-834 #label INA
#cross-references EMBL:S79463; NID:g110598; PID:g110599
```

```
ACCESSIONS S66498
REFERENCE S66498
#authors Inagaki, S.; Furuyama, T.; Iwahashi, Y.
#journal FEBS Lett. (1995) 370:269-272
#title Identification of a member of mouse semaphorin family.
#cross-references MUID:95385809
#accession S66498
#status preliminary
#molecule_type mRNA
#residues 1-834 #label INA
#cross-references EMBL:S79463; NID:g110598; PID:g110599
```


THIS PAGE BLANK (USPTO)

W P E R L A (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:29:29 1999; Maspar time 18.79 Seconds

Tabular output not generated. 904.119 Million cell updates/sec

Title: >US-09-041-236-2

Description: (1-601) from US09041236.pep (36 of 45)

Perfect Score: 4488

Sequence: 1 LLLLLWAAASAGHLRSGP.....EAQGSYFRAQHQLLPED 601

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 50.410; Variance 76.174; scale 0.662

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	470	10.5	1 VA39_VACCC	PROTEIN A39.	3.64e-86
2	463	10.3	441 1 VA39_VACCV	PROTEIN A39.	2.31e-84
3	163	3.6	1871 1 SEX_HUMAN	TRANSMEMBRANE PROTEIN	1.04e-12
4	105	2.3	227 1 US08_HCMVA	HYPOTHETICAL PROTEIN H	4.13e-02
5	102	2.3	275 1 NRT2_RAT	T-CELL ECTO-ADP-RIBOSY	1.21e-01
6	99	2.2	275 1 NRT1_RAT	T-CELL ECTO-ADP-RIBOSY	3.43e-01
7	99	2.2	460 1 EF1A_TRIRE	ELONGATION FACTOR 1-AL	3.43e-01
8	99	2.2	591 1 VRP2_SALTU	65 KD VIRULENCE PROTEI	3.43e-01
9	99	2.2	591 1 VRP2_SALCH	65 KD VIRULENCE PROTEI	3.43e-01
10	99	2.2	591 1 VRP2_SALEN	65 KD VIRULENCE PROTEI	3.43e-01
11	99	2.2	593 1 VRP2_SALDU	65 KD VIRULENCE PROTEI	3.43e-01
12	99	2.2	916 1 PHA1_AJECA	PLASMA MEMBRANE ATPASE	3.43e-01
13	98	2.2	1132 1 DNBI_HSV6U	MAJOR DNA-BINDING PROT	4.83e-01
14	95	2.1	255 1 HA21_HUMAN	HLA CLASS II HISTOCOMP	1.32e+00
15	94	2.1	329 1 RBSR_ECOLI	RIBOSE OPERON REPRESSO	1.84e+00
16	93	2.1	380 1 DP3B_MYCPN	DNA POLYMERASE III, BE	2.55e+00
17	95	2.1	518 1 VL2_HPV5B	MINOR CAPSID PROTEIN L	1.32e+00
18	95	2.1	518 1 VL2_HPV05	MINOR CAPSID PROTEIN L	1.32e+00
19	95	2.1	683 1 AMOH_ARTGO	HISTAMINE OXIDASE (EC	1.32e+00
20	93	2.1	827 1 Y223_METJA	HYPOTHETICAL PROTEIN M	2.55e+00
21	94	2.1	947 1 PNA2_YEAST	PLASMA MEMBRANE ATPASE	1.84e+00
22	94	2.1	959 1 NI00_YEAST	NUCLEOPORIN NUP100/NSP	1.84e+00
23	94	2.1	982 1 POL_HTLV2	POL POLYPROTEIN [COMTA	1.84e+00

24	93	2.1	1029	1	END1_YEAST	VACUOLAR BIOGENESIS PR	2.55e+00
25	94	2.1	1069	1	ENTK_MOUSE	ENTEROPEPTIDASE (EC 3.	1.84e+00
26	94	2.1	1115	1	IREL_YEAST	SERINE/THREONINE-PROTE	1.84e+00
27	94	2.1	1200	1	DDX8_CAEEL	PUTATIVE PRE-MRNA SPLI	1.84e+00
28	95	2.1	1390	1	MET_HUMAN	HEPATOCYTE GROWTH FACT	1.32e+00
29	90	2.0	119	1	RL19_MYCPN	50S RIBOSOMAL PROTEIN	6.63e+00
30	90	2.0	213	1	AMEX_BOVIN	AMELOGENIN, CLASS I PR	6.63e+00
31	90	2.0	254	1	HA22_HUMAN	HLA CLASS II HISTOCOMP	6.63e+00
32	91	2.0	348	1	DDL_ENTFA	D-ALANINE--D-ALANINE L	4.84e+00
33	91	2.0	356	1	VP39_NPVLD	MAJOR CAPSID PROTEIN.	4.84e+00
34	90	2.0	380	1	GBB_MAI2E	GUANINE NUCLEOTIDE-BIN	6.63e+00
35	92	2.0	442	1	CEM1_YEAST	3-OXOACYL-[ACVL-CARRIE	3.52e+00
36	92	2.0	460	1	EF1A_NEUCR	ELONGATION FACTOR 1-AL	3.52e+00
37	91	2.0	498	1	ELAS_PSAE	PSEUDOLYSIN PRECURSOR	4.84e+00
38	91	2.0	598	1	NOT_HUMAN	IMMEDIATE-EARLY RESPON	4.84e+00
39	89	2.0	598	1	NURI_MOUSE	PROBABLE NUCLEAR HORMO	9.06e+00
40	91	2.0	707	1	GCVK_HCMVA	GAMCICLOVIR KINASE (EC	4.84e+00
41	89	2.0	966	1	AMPN_HUMAN	AMINOPEPTIDASE N (EC 3	9.06e+00
42	92	2.0	1132	1	DNBI_HSV6Z	MAJOR DNA-BINDING PROT	3.52e+00
43	92	2.0	1663	1	CO3_MOUSE	COMPLEMENT C3 PRECURSO	3.52e+00
44	91	2.0	1666	1	CO3_CAVPO	COMPLEMENT C3 PRECURSO	4.84e+00
45	90	2.0	1874	1	POLR_KYMWJ	RNA REPLICASE POLYPROT	6.63e+00

ALIGNMENTS

RESULT 1
ID VA39_VACCC STANDARD; PRT; 403 AA.
AC P21062;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE PROTEIN A39.
GN A39R.
OS VACCINIA VIRUS (STRAIN COPENHAGEN).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91021027.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RT "The complete DNA sequence of vaccinia virus.";
RL VIROLOGY 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RL VIROLOGY 179:517-563(1990).
CC -----
CC This swiss-prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M35027; G335517; -
DR PIR; E42521; E42521.
SQ SEQUENCE 403 AA; 45741 MW; FE0DB9AB CRC32;

Query Match 10.5%; Score 470; DB 1; Length 403;
Best Local Similarity 33.5%; Pred. No. 3.64e-86;
Matches 86; Conservative 59; Mismatches 95; Indels 17; Gaps 16;

Db	77	LVCGTNGNPKWK-IDGSDDPKRRGRGYPYQNSKVTIISYN-ECVLSDINISK-SG-I	132
Qy	109	LACGTNARHPSCWNLVNGTVVPLGEMRGYAPFSPDENSELVLFEGDEVYSTIRKQYNGKI	168
Db	133	KWRREDPCGDIYLTADNVIPIKDG-LRGAFAVDKGDYD-KVYILFTDTIGSKR--I-VK	187
Qy	169	PRFRIRIGE-S-EDYTSDTVQNQPFKATIVHODQYDDKIYFFFDNPDNPKRNPAPLN	226

QY 131 LGE-MRGYAPFSPDENSLSVLFEGDEVYSTIRKOEYNGK-IPRFRIRGESE-LYTSDTVM 187
Db 252 ES 253
QY 188 QN 189

RESULT 6
ID NRT1_RAT STANDARD; PRT; 275 AA.
AC P17982;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 PRECURSOR (EC 2.4.2.31) (T-CELL
DE NAD(P)(+)-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL MONO(ADP-
DE RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN
DE RT6.1).
GN ART2A OR RT6-A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEWIS A.
RX MEDLINE: 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell
RT alloantigen RT6.1".
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
RN [2]
RP MUTAGENESIS OF GLN-207.
RX MAEHAMA T., HOSHINO S.-I., KATADA T.;
RA "Increase in ADP-ribosyltransferase activity of rat T lymphocyte
RT alloantigen RT6.1 by a single amino acid mutation".
RT FEBS LETT. 388:189-191(1996).
CC -!- FUNCTION: HAS NAD+ GLYCOPOLYMERASE ACTIVITY AND EXTREMELY LOW ADP-
CC RIBOSYLTRANSFERASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
CC N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -!- CATALYTIC ACTIVITY: NAD(+) + H(2)O = NICOTINAMIDE + ADP-RIBOSE.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -!- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL; X52082; G57168; -.
CC DR EMBL; M31138; G206804; -.
CC DR PIR; S08464; S08464.
CC DR PROSITE; PS01291; ART; 1.
CC DR PFAM; PF01129; ART; 1.
CC KW TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
CC T-CELL DIFFERENTIATION; GPI-ANCHOR.
CC FT SIGNAL 1 20
CC FT CHAIN 21 246 T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
CC FT PROPEP 247 275 HYDROPHOBIC. REMOVED DURING MATURATION
CC (BY SIMILARITY).
CC FT LIPID 246 246 GPI-ANCHOR (BY SIMILARITY).
CC FT ACT_SITE 209 209 BY SIMILARITY.
CC FT CARBOHYD 58 58 POTENTIAL.
CC FT MUTAGEN 207 207 Q->E: INCREASED ADP-RIBOSYLTRANSFERASE
CC ACTIVITY.
CC SEQUENCE 275 AA; 31388 MW; 523B1A84 CRC32;

Query Match 2.2%; Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 3.43e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;
Db 194 LGVYIKFS-FYPDQEE-VLIPGYEVYQKVRVQGYNEIFLDSFKRKKSNLYSSAGTR 251
QY 131 LGE-MRGYAPFSPDENSLSVLFEGDEVYSTIRKOEYNGK-IPRFRIRGESE-LYTSDTVM 187
Db 252 ES 253
QY 188 QN 189

RESULT 7
ID EF1A_TIRE STANDARD; PRT; 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
DE TEFL.
GN TEFL.
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-QM9414 / RUTC-30;
RX MEDLINE: 94124021.
RA NAKARI T., ALATALO E., PENTTILA M.;
RT "Isolation of Trichoderma reesei genes highly expressed on glucose-
RT containing media: characterization of the tefl gene encoding
RT translation elongation factor 1 alpha".
RL GENE 136:313-318(1993).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL; Z23012; G312887; -.
CC DR PIR; S35772; S35772.
CC DR PROSITE; PS00301; EFATOR_GTP; 1.
CC DR PFAM; PF00009; GTP_EFTU; 1.
CC DR HSP; P07157; IATP; 1.
CC KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
CC FT NP_BIND 15 22 GTP (BY SIMILARITY).
CC FT NP_BIND 92 96 GTP (BY SIMILARITY).
CC FT NP_BIND 154 157 GTP (BY SIMILARITY).
CC SQ SEQUENCE 460 AA; 49830 MW; B9ABAB2 CRC32;

Query Match 2.2%; Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 3.43e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;
Db 210 YKGWKEIKAGFTG-KTLLAIDSTIEPKR-PTDKPLRLPLQDV 252
QY 478 YCGWDOGRCSIIYSYRSVLSQINPAEPKPCPNKPKAPLQKV 522

RESULT 8
ID VRP2_SALTY STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN MKAA.
OS SALMONELLA TYPHIMURIUM.
OG PLASMID 96 KB VIRULENCE PEX102.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
ON SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TML R66;
RX MEDLINE; 90136009.
RA TAIRA S., RHEN M.;
RT "Identification and genetic analysis of mkaA -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth.";
RL MICROB. PATHOG. 7:165-173(1989).
RN [2]
RP SEQUENCE OF 1-10.
RX MEDLINE; 91244158.
RA TAIRA S., BAUMANN M., RIIKONEN P., SUKUPOLVI S., RHEN M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium.";
RL FEMS MICROBIOL. LETT. 61:319-323(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLA.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; Z15042; G47783; -.
DR PIR; A54540; A54540.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65440 MW; B69F25C3 CRC32;
Query Match 2.2%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.43e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLGKTAARLSDPOAASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVNLN 219
QY 92 RCENYITLLERSEGLACGTNARHPSCNVLNGTVPLGEMRGYAPSPDENSILVFE 151
Db 220 GNEAGDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVSTIRKQYNGKIPFRIRGESELY 181
RESULT 9
ID VRP2_SALCH STANDARD; PRT; 591 AA.
AC P17450;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN (PROTEIN M2, IN MBA REGION).
OS SALMONELLA CHOLERAE-SUIS (SALMONELLA-ENTERICA).
OG PLASMID PKDCS0.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
ON SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RF-1;
RX MEDLINE; 90245675.
RA MATSUI H.;
RT "Nucleotide sequences of genes encoding 32 kDa and 70 kDa
RT polypeptides in mba region of the virulence plasmid, pKDCs0, of

RT Salmonella choleraesuis.";
RL NUCLEIC ACIDS RES. 18:2181-2181(1990).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLA.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; X52035; G46898; -.
DR PIR; S09498; S09498.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65341 MW; BE4A439A CRC32;
Query Match 2.2%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.43e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLGKTAARLSDPOAASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVNLN 219
QY 92 RCENYITLLERSEGLACGTNARHPSCNVLNGTVPLGEMRGYAPSPDENSILVFE 151
Db 220 GNEAGDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVSTIRKQYNGKIPFRIRGESELY 181
RESULT 10
ID VRP2_SALEN STANDARD; PRT; 591 AA.
AC P55220;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN SPVB.
OS SALMONELLA ENTERITIDIS.
OG PLASMID PNL2001.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
ON SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AL1190;
RX MEDLINE; 94362897.
RA SUZUKI S., KOMASE K., MATSUI H., ABE A., KAWAHARA K., TAMURA Y.,
RA KIJIMA M., DANBARA H., NAKAMURA M., SATO S.;
RT "Virulence region of plasmid PNL2001 of Salmonella enteritidis.";
RL MICROBIOLOGY 140:1307-1318(1994).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLA.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; D14490; G517164; -.
DR PIR; S09498; S09498.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65350 MW; 9C5B2EE3 CRC32;
Query Match 2.2%; Score 99; DB 1; Length 591;

```
Best Local Similarity 23.3%; Pred. No. 3.43e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPAQASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 92 RDCENYITLLRRSGLACGTNARHPSCNVLNGTVVPLGEMRGYAPFSPDENSLLVFE 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 GNEAGRDRSAMRYSLSKV-OYGNATPAADLY 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 152 GDEVSTIRKQYNGKIPFRFRIRGESELY 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
ID VRP2_SALDU STANDARD; PRT; 593 AA.
AC P24419;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN VSDC.
OS SALMONELLA DUBLIN.
OC PLASMID PSDL2.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91251759.
RA KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINEY D.;
RT "Molecular analysis of the virulence locus of the Salmonella dublin
   plasmid pSDL2.";
RL MOL. MICROBIOL. 5:307-316(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L07305; G409249;
DR PROSITE; PS00134; ATPASE_L1_E2; 1.
DR PFAM; PF00122; E1-E2 ATPase; 1.
KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
KW ATP-BINDING.
FT DOMAIN 1 111 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 112 134 1 (POTENTIAL).
FT DOMAIN 135 136 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 137 155 2 (POTENTIAL).
FT DOMAIN 156 279 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 280 305 3 (POTENTIAL).
FT DOMAIN 306 320 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 321 350 4 (POTENTIAL).
FT DOMAIN 351 711 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 712 734 5 (POTENTIAL).
FT DOMAIN 735 754 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 755 776 6 (POTENTIAL).
FT DOMAIN 777 817 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 818 843 7 (POTENTIAL).
FT DOMAIN 844 874 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 875 916 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 916 916 ASP/GLU-RICH (ACIDIC).
FT MOD_RES 374 374 PHOSPHORYLATION (BY SIMILARITY).
FT BINDING 470 470 ATP (BY SIMILARITY).
SQ SEQUENCE 916 AA; 98884 MW; 4A468A44 CRC32;

Query Match 2.2%; Score 99; DB 1; Length 916;
Best Local Similarity 37.2%; Pred. No. 3.43e-01;
Matches 16; Conservative 8; Mismatches 16; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPAQASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 92 RDCENYITLLRRSGLACGTNARHPSCNVLNGTVVPLGEMRGYAPFSPDENSLLVFE 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 GNEAGRDRSAMRYSLSKV-OYGNATPAADLY 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 152 GDEVSTIRKQYNGKIPFRFRIRGESELY 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
ID PMAL_AJECA STANDARD; PRT; 916 AA.
AC Q07421;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).
GN PMA...
OS AJELLOWYCES CAPSULATA (HISTOPLASMA CAPSULATUM).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
```

```

RN SEQUENCE FROM N.A.
RX MEDLINE: 95266321.
RA GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA MARTIN M.E., EFSTATHIOU S., CRAYTON M., MACAULAY H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution.";
RL VIROLOGY 209:29-51(1995).
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES' DNA-BINDING PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; X92436; G1044871; -.
CC EMBL; X83413; G854020; -.
CC DR PFAM; PF00747; Viral_DNA_bp; 1.
CC KW DNA-BINDING; DNA REPLICATION; ZINC-FINGER; NUCLEAR PROTEIN.
CC FT ZN_FING 459 475 C4-TYPE.
CC SQ SEQUENCE 1132 AA; 127762 MW; D8699A02 CRC32;
Query Match 2.2%; Score 98; DB 1; Length 1132;
Best Local Similarity 28.2%; Pred. No. 4.83e-01;
Matches 24; Conservative 18; Mismatches 37; Indels 6; Gaps 6;
Db 435 WNLNRMVYVGNAGNHTI-YNHLVNCNANICEFCGDKC-C-QSCIGTAMVVGTRLPAP 491
QY 297 WNSYAVCVISGDKDKVFTSSLGXGYSLSLPNRPGRKCLPDQDPPIPTETQVADRHEVA 356
Db 492 KNYK-KEPLVSMF-SRY-YAEVDI 513
QY 357 QRVEPMGLTPLFHSKHYQKVAV 381
RESULT 14
ID HA21_HUMAN STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84168117.
RA AUFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPES D.,
RA STROMINGER J.L.;
RT "Isotypic and allotypic variation of human class II
RT histocompatibility antigen alpha-chain genes.";
RL NATURE 308:327-333(1984).
RN [2]
RP SEQUENCE OF 40-255 FROM N.A.
RX MEDLINE: 83065171.
RA AUFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT "cDNA clone for the heavy chain of the human B cell alloantigen DC1:
RT strong sequence homology to the HLA-DR heavy chain.";
RL PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; J00199; -. NOT_ANNOTATED_CDS.
CC PIR; A02212; HLHUCD.
CC DR PROSITE; PS00290; IG_MHC; 1.
CC PFAM; PF00047; Ig_1.
CC DR PFAM; PF00993; MHC_II_alpha; 1.
CC DR HSP; P01910; IIAK.
CC KW MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
CC FT SIGNAL 1 23
CC FT CHAIN 24 255 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
CC DQ(1) ALPHA CHAIN.
CC FT DOMAIN 24 110 EXTRACELLULAR ALPHA-1.
CC FT DOMAIN 111 204 EXTRACELLULAR ALPHA-2.
CC FT DOMAIN 205 217 CONNECTING PEPTIDE.
CC FT TRANSMEM 218 240
CC FT DOMAIN 241 255 CYTOPLASMIC TAIL.
CC FT DISULFID 133 189 BY SIMILARITY.
CC FT CARBOHYD 104 104 POTENTIAL.
CC FT CARBOHYD 144 144 POTENTIAL.
CC SQ SEQUENCE 255 AA; 28105 MW; 5C46DD59 CRC32;
Query Match 2.1%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1.32e+00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;
Db 25 DIVADSVQAQLG-VNLXSYSGSQYSHEFDGDEEYVDLERKETVMOLPLFRFR 78
QY 122 NLVNGTVPPLGMRGVAPSPDENSELSLVFEGDEV-YSTIRKQYNGKIPRRIR 175
RESULT 15
ID RBSR_ECOLI STANDARD; PRT; 329 AA.
AC P25551;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOSE OPERON REPRESSOR.
GN RBSR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 1-6 AND 250-256.
RX STRAIN-K12;
RX MEDLINE: 93278299.
RA MAUZY C.A., HERMOSON M.A.;
RT "Structural and functional analyses of the repressor, RbsR, of the
RT ribose operon of Escherichia coli.";
RL PROTEIN SCI. 1:831-842(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE: 93315143.
RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL GENOMICS 16:551-561(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RA MAUZY C.A., HERMOSON M.A.;
RT "Structural homology between rbs repressor and ribose binding protein
RT implies functional similarity.";
RL PROTEIN SCI. 1:843-849(1992).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RBSDACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING
```

```
CC THE RBS OPERON TRANSCRIPTIONAL START SITE. THE AFFINITY FOR THE
CC RBS OPERATOR IS REDUCED BY ADDITION OF RIBOSE, CONSISTENT WITH
CC RIBOSE BEING THE INDUCER OF THE OPERON.
CC -!- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13169; G147517; -.
DR EMBL; L10328; G290603; ALT_INIT.
DR EMBL; AE000452; G1790194; -.
DR EMBL; D10466; G471110; -.
DR PIR; A41828; A41828.
DR ECOGENE; EG10819; RBSR.
DR PROSITE; PS00356; HTH_LACI_FAMILY; 1.
DR PFAM; PF00356; lacI; 1.
DR PFAM; PF00532; peripla_BP_like; 1.
DR HSP; P15039; IPRV.
DR TRANSCRIPTION REGULATION; REPRESSOR; DNA-BINDING.
FT INIT_NET 0 0
FT DNA_BIND 3 22 H-T-H MOTIF (BY SIMILARITY).
FT CONFLICT 11 12 GV->L (IN REF. 1).
SQ SEQUENCE 329 AA; 36480 MW; 424E72A6 CRC32;

Query Match 2.1%; Score 94; DB 1; Length 329;
Best Local Similarity 34.6%; Pred. No. 1.84e+00;
Matches 18; Conservative 14; Mismatches 14; Indels 6; Gaps 6;

Db 109 "LMQKRVGGLLLC-TETHQPS-REIMQRYTPV-TVMMDWAPFDGSD-LI 156
Oy 99 "LLERSEGLA-CGTNARHPSCWNLVNG-TVVPLGEMRGYAPFSPDENS LV 148
```

Search completed: Thu Jul 8 19:29:53 1999
Job time : 24 secs.

Db	333	AVCVYSLGIDIKVFTRSSLUKGVHSSLPNPRGCKLPDQQPIPTETFOVADRHPEVAQRVE	392
Qy	301	AVCVYSLGIDIKVFTRSSLUKGVHSSLPNPRGCKLPDQQPIPTETFOVADRHPEVAQRVE	360
Db	393	PMGPLKTPLFHSHKYHYOKVAVHRMOASHGETFHVLYLTTORGTHIKVPEGEQHSFAFN	452
Qy	361	PMGPLKTPLFHSHKYHYOKVAVHRMOASHGETFHVLYLTTORGTHIKVPEGEQHSFAFN	420
Db	453	IMEIQPFRRAAAIQTMSLDAERRKLYVSSQEVSVQPLDLCVYGGCGHCLMSRDPCG	512
Qy	421	IMEIQPFRRAAAIQTMSLDAERRKLYVSSQEVSVQPLDLCVYGGCGHCLMSRDPCG	480
Db	513	WDGRCISYISYSSRSVQSINPAEPHKKCPNPKDPKAPLOKVSILAPNSRYILSCPMESSRH	572
Qy	481	WDGRCISYISYSSRSVQSINPAEPHKKCPNPKDPKAPLOKVSILAPNSRYILSCPMESSRH	540
Db	573	ATYSWRHKENVQSCPEGHOSPNCLIFTENLTAQOYGHYFCEAQEGSYFREAQHWOLLPE	632
Qy	541	ATYSWRHKENVQSCPEGHOSPNCLIFTENLTAQOYGHYFCEAQEGSYFREAQHWOLLPE	600
Db	633	D 633	
		I	
Qy	601	D 601	
RESULT 2			
ID	088371	PRELIMINARY; PRT; 393 AA.	
AC	088371;		
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	SEMAPHORIN L (FRAGMENT).		
GN	SEMAL.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE: 98389619.		
RT	LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;		
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA		
RL	viruses."		
RL	GENOMICS 51:340-350(1998).		
DR	ENBL; AF030699; G3523117; -.		
FT	NON_TER 393 393		
SQ	SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;		
Query Match 53.4%; Score 2398; DB 11; Length 393;			
Best Local Similarity 90.4%; Pred. No. 0.00e+00;			
Matches 329; Conservative 16; Mismatches 15; Indels 4; Gaps 2			
Db	33	LLLVFWAAASAQGHSRSGPRISAVWKG--QDHVDFSQPEPHTVLFHEPGSFSVMVWGR	89
Qy	1	LLLLLWAAASAQGHLRSGPRIFAWKGHVQDRVDFGQTEPTVLFHEPGSSVVMVWGR	60
Db	90	GKVIYHNFPEGNASRVTVNGSTGSKQDQKDCGNVITLLRRGNGLVCGTNAKRPSC	149
Qy	61	GKVIYDFDFPEGNASRVTVNGSTGSKLDKRCDCNVITLLRRSEGLACGTNARHPSC	120
Db	150	WNLVNDSVYMSLGMKGVAPTSFSPENDSLVLFEGDEVYSTIRKQYNGKIPRFRIRIGESE	209
Qy	121	WNLVNGTVV-PLGEMRGVAPTSFSPENDSLVLFEGDEVYSTIRKQYNGKIPRFRIRIGESE	179
Db	210	LYTSDTVMQNPOFIKATIVHQDQAYDDKIYFFREDNPKNPEAPLVNSRVAQLCRGDQG	269
Qy	180	LYTSDTVMQNPOFIKATIVHQDQAYDDKIYFFREDNPKNPEAPLVNSRVAQLCRGDQG	239
Db	270	GESSLSVSKWNTFLKAMLYCSDAATNRNPNLDVLLPDPSQWRDTRVYGVFSPNPNY	329
Qy	240	GESSLSVSKWNTFLKAMLYCSDAATNKNPNLDVLLPDPSQWRDTRVYGVFSPNPNY	299
Db	330	SAYCVYSLGIDIRVFTSSLAGYHMGLSNPRPGMKLPKQPIPTETFOVADSHPEVAQRV	389

Qy	300	SAVCVYSLGDDIDKVFRTSSLUKGHSSLNRPKCLPDDQQPIPTFTFQVADRHPEVAQRV	359
Dd	390	EPMG 393 	
Qy	360	EPMG 363	
RESULT 3			
ID	Q64906	PRELIMINARY; PRT; 653 AA.	
AC	Q64906;		
DT	01-NOV-1996 (TREMBLER. 01, CREATED)		
DT	01-NOV-1996 (TREMBLER. 01, LAST SEQUENCE UPDATE)		
DE	01-NOV-1998 (TREMBLER. 08, LAST ANNOTATION UPDATE)		
DE	DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.		
OS	ALCELAAPHINE HERPESVIRUS 1.		
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;		
OC	GAMMAHERPESVIRINAE.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C500;		
RX	MEDLINE; 97201573.		
RR	ENSSER A., FLECKENSTEIN B.;		
RT	"Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";		
RL	J GEN. VIROL. 76:1063-1067(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C500;		
RX	MEDLINE; 97404659.		
RR	ENSSER A., PFLANZ R., FLECKENSTEIN B.;		
RT	"Primary structure of the alcelaphine herpesvirus 1 genome.";		
RL	J. VIROL. 71:6517-6525(1997).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C500;		
RR	ENSSER A., PFLANZ R., FLECKENSTEIN B.;		
RL	SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.		
DR	EMBL; U18243; G1000717; -.		
DR	EMBL; AF005370; G2337970; -.		
SQ	SEQUENCE 653 AA; 73645 MW; CCCC53C9 CRC32;		
Query Match 42.0%; Score 1883; DB 14; Length 653;			
Best Local Similarity 46.3%; Pred. No. 0.00e+00;			
Matches 274; Conservative 105; Mismatches 200; Indels 13; Gaps 13;			
Dd	57	LLML-L-SAITAAKSRFDTKPRLLIVNLTDGFQGHRF-FGPQPHTVLVFLHSLNSSDYYVGNN 114	
Qy	1	LLLLLWAAAAAQAAGLRSGPRIFAVNKHGVQDRVDFGQTETHTLVLFHEPGSSVVWGGR 60	
Dd	115	NTIYDFDAHSSNASATALINITSTHNRSLSSTCENFIILLHNQTOGLLACGTNSOKPSC 174	
Qy	61	GKYVDFDPGGKNASVRTVMIGTSGCDLKDCDENYITLLERRSEGLLACGTNARHPSC 120	
Dd	175	W-LINNLTTFQLPKGLAPFPSSGNLVLFDOONDYTSINLYKSLSGS-HKFERRIAGOV 232	
Qy	121	WNLVNGVIYP-LGEMRGAPFSPDENSILVLFEGDEVYSIRK-QEYNGKIPIRFRIRGES 178	
Dd	233	ELYTSDTAHRPOFQVATAVHKNESYDDDKTIYFFQBNSHSDFKQFPHTVPRGVQCSSDQ 292	
Qy	179	ELYTSDTMQNOPFIKATIVHQDAQYDDDKIIYFFREDNPDKNPEAPLNVSRVAQLCRGDQ 238	
Dd	293	GGSSLSVYKWTTLFKARLACVDYDGRINYELQDIFIWAQPNSEETLLIYGLFLSPKN 352	
Qy	239	GGSSLSVSKWNTFLKAWLVCSAAANKNFNRLODVFLLPDPSGQWRDTRVYGVSFNPNW 298	
Dd	353	FSAVCVFTVWDIDHVFRTSKLNYHHKLPTPRPGQCMKNQHVPTEFTTFQVADRYPEVADP 412	
Qy	299	YSACVYSLGDDIDKVFRTSSLUKGHSSLNRPKCLPDDQQPIPTFTFQVADRHPEVAQR 358	
Dd	413	VYQKNAMPFIIOISKYIYTKLLVRYVEYG-GVFPAIFYLTTIKTGITHIYRVYEDSNSTT 471	
Qy	359	VEPMGPLKTPLFFSKYHYQKVAVRHQASHGETFH-VLXYLTTDRGTTHIKHWEPGEQEHSF 417	

AC 08632;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV ISOFORM B.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ECKHARDT F., MEYERHANS A.;
RT "molecular cloning and expression pattern of a murine semaphorin
RT homologous to H-sema IV";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF080090; G337766; -;
SQ SEQUENCE 785 AA; 88493 MW; 6FFCD8F5 CRC32;

Query Match 15.2%; Score 682; DB 11; Length 785;
Best Local Similarity 31.8%; Pred. No. 1.51e-140;
Matches 141; Conservative 119; Mismatches 140; Indels 43; Gaps 31;

Db 202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRTLKGTAMRTDQYNSRWLNDSFTH 260
QY 139 PFPDENSELVLFEGDEVYSTIRKQYNGK-IPFRIRIRGESELYTS--DTV-MQNPOFIK 194
Db 261 AELIPDSAEKNDKLYFFFRERSAE-APONPAYARIGRICLNDDGGHCL-VNKHSTFL 318
QY 195 ATIVHOD-QAYDDKIYFFREDNPKNPEAPLNVSRVAQLCRDQGGESSLSVSKWNTFL 253
Db 319 KARLVCSVPGEIGIETHFDELQDVV-OQTODI-RNPVIYAVFTSSGVSFRGSVAVCVISM 376
QY 254 KAMLVCS---DAATNKNFNRLQDVLLPDPSCQWRDTRYGVFSNP-WNY--SAVCVISL 307
Db 377 ADIRVFNFGPFAHKEGPNYQWMPFSGKMPYPRGTCPGGTFTPSMKSTKDYDPEVINEMR 436
QY 308 GIDKIVERTS-SLK-G-YH---SS-LPNRPKCLPDQ-QP-I-PTETF--QVAD--R 351
Db 437 THPLMYQAVYPLORPLVVRT-GAPYRLTTIAVDQVDAADG-RYEVFLGTRGIVOKVI 494
QY 352 -HPEVAQRVEPMG--PLKTPLFHSYHYQKVAHVRMQASHGETFHYLYLTDRGTIHKVY 408
Db 495 VLPKDDQEELEMLEVEVEKPAVKTMTISSKROOLYVASAVGVTLSLHRCQAYGAA 554
QY 409 E-PGGEHSEFAFNIQIFRRAAIIQTLMSDAERKLYVSSQWVQVPLDCEYVGG 467
Db 555 CADCCCLARDPYCAWDQACSRYSATSKRRSRQDVHRGNPIRCQGRFNSNANKNAVESVQ 614
QY 468 CHGCLMSRDPYCGWDQGRGIS-IYSSE-SVLQSIINPAEPHKECP--NPKPKAPLQKVS 523

Db 615 YGVAGSAFLEQCPSPQATVWK 637
QY 524 LA-PNSRYLSCPMESRHATYSW 545

RESULT 14
ID Q13275 PRELIMINARY; PRT: 785 AA.
AC Q13275; Q13274;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA NELSON J., BIEWALD T.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 394-436 FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE; 96210603.
RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,

RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT cancer deletion region and demonstrate distinct expression
RT patterns";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR EMBL; AC000063; G1669373; -;
DR EMBL; U32172; G955788; -;
DR EMBL; U32171; G955786; -;
DR PFAM; PF00047; Ig 1.
SQ SEQUENCE 785 AA; 88381 MW; CD175765 CRC32;

Query Match 15.2%; Score 681; DB 4; Length 785;
Best Local Similarity 31.4%; Pred. No. 2.78e-140;
Matches 139; Conservative 122; Mismatches 139; Indels 43; Gaps 31;

Db 202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRTLKGTAMRTDQYNSRWLNDSFTH 260
QY 139 PFPDENSELVLFEGDEVYSTIRKQYNGK-IPFRIRIRGESELYTS--DTV-MQNPOFIK 194
Db 261 AELIPDSAEKNDKLYFFFRERSAE-APQSPAYARIGRICLNDDGGHCL-VNKHSTFL 318
QY 195 ATIVHOD-QAYDDKIYFFREDNPKNPEAPLNVSRVAQLCRDQGGESSLSVSKWNTFL 253
Db 319 KARLVCSVPGEIGIETHFDELQDVV-OQTODV-RNPVIYAVFTSSGVSFRGSVAVCVISM 376
QY 254 KAMLVCS---DAATNKNFNRLQDVLLPDPSCQWRDTRYGVFSNP-WNY--SAVCVISL 307
Db 377 ADIRVFNFGPFAHKEGPNYQWMPFSGKMPYPRGTCPGGTFTPSMKSTKDYDPEVINEMR 436
QY 308 GIDKIVERTS-SLK-G-YH---SS-LPNRPKCLPDQ-QP-I-PTETF--QVAD--R 351
Db 437 SHPLMYQAVYPLORPLVVRT-GAPYRLTTIAVDQVDAADG-RYEVFLGTRGIVOKVI 494
QY 352 -HPEVAQRVEPMG--PLKTPLFHSYHYQKVAHVRMQASHGETFHYLYLTDRGTIHKVY 408
Db 495 VLPKDDQEELEMLEVEVEKPAVKTMTISSKROOLYVASAVGVTLSLHRCQAYGAA 554
QY 409 EPGGEHSEFAFNIQIFRRAAIIQTLMSDAERKLYVSSQWVQVPLDCEYVGG 467
Db 555 CADCCCLARDPYCAWDQACSRYSATSKRRSRQDVHRGNPIRCQGRFNSNANKNAVESVQ 614
QY 468 CHGCLMSRDPYCGWDQGRGIS-IYSSE-SVLQSIINPAEPHKECP--NPKPKAPLQKVS 523

Db 615 YGVAGSAFLEQCPSPQATVWK 637
QY 524 LA-PNSRYLSCPMESRHATYSW 545

RESULT 15
ID Q15704 PRELIMINARY; PRT: 785 AA.
AC Q15704;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE; 96226360.
RA ROCHE J., BOLDOG F., ROBINSON M., ROBINSON L., VARELLA-GARCIA L.,
RA SWANTON M., WAGGONER B., FISHEL R., FRANKLIN W., GEMMILL R.,
RA DRABKIN H.;
RT "Distinct 3p21.3 deletions in lung cancer and identification of a new
RT human semaphorin";
RL ONCOGENE 12:1289-1297(1996).
DR EMBL; U33920; G1000207; -;
DR PFAM; PF00047; Ig 1.
SQ SEQUENCE 785 AA; 88385 MW; 9971F14E CRC32;

Query Match 15.1%; Score 679; DB 4; Length 785;

Rest Local Similarity 31.6%;	Pred. No. 9.34e-140;	
Matches 140;	Conservative 121;	Mismatches 139;
	Indels 43;	Gaps 31;
Db	202	PYDEKLTASALINEELYAGVYI--DFMGTDAAITFTLGKTMARTQYNSRWLNDPSFH 260
		: : : : : : : : : : : : : : : : : : : : : : : :
Qy	139	PFPDENSLVLFGEDEVYSIRKQEYNGK-IPFERRIRGESELYTS--DTV-MQNQFIK 194
		: : : : : : : : : : : : : : : : : : : : : : : :
Db	261	AELIPDSAEARNDDKLYFFFRERSAE-AQSPAVYAVIRIGICLNDGDGHCOL-YNKWSTEL 318
		: : : : : : : : : : : : :
Qy	195	ATIVHQD-QAYDGTIIYFFREDNPDKNPEAPLNVSRVAQLCRDQGGESSLVSXKNTEL 253
		: : : : : : : : : : : : :
Db	319	KARLVCSVPGEDGTETHFDELDQVYV-QQTQDV-RNPVIYATISGGSVFRGSACVYSM 376
		: : : : : : : : : : : : : : : : : : : : : : :
Qy	254	KAMLVCS---DAATNKNFNQLQVFLPDPDSGQWRDTRYGVESNP-WNY--SAVCVYSL 307
		: : : : : : : : : : : : : : : : : : : : : : : :
Db	377	ADIRWVNGFPAHKEGNYQWMPFSGKMPYPRGTCPGGTFSMKSTKDYDPDEVNMR 436
		: : : : : : : : : : : : : : : : : : : : : : : :
Qy	308	GDIDKVERTS-SLK-G-YH-----SS-LPNPRGKCLPDQ-QP-I-PTETF--QVAD--R 351
		: : : : : : : : : : : : : : : : : : : : : : : :
Db	437	SHPLMYQAVYPLQRRPLVRT-GAPYRLTTIAVDQVDAGDG-RYEVLFGTDRGTQKYI 494
		: : : : : : : : : : : : : : : : : : : : : : : :
Qy	352	-HPEVAQRVEPMG--PLKTLFHSKYHYQKVAHVHRMQASHGETFHVLYLTTRDGTIHKVV 408
		: : : : : : : : : : : : : : : : : : : : : : : :
Db	495	VLPKDDQEMEELMEEVEFKDAPVKTMTIISKROOLYVASAVGTHLSLHRCQAYGAA 554
		: : : : : : : : : : : : : : : : : : : : : : : :
Qy	409	E-PGEQHSFAFNIMEIQPFRAAAIQTMSLDERRKLYVSSQEWVSQVPLDCEYVGG 467
		: : : : : : : : : : : : : : : : : : : : : : : :
Db	555	CADCCCLARDPYCAWDGQACSRYTASSKRRSRRRQDVRRGNPIRQCRGFNSNANKNAVESYQ 614
		: : : : : : : : : : : : :
Qy	468	CHGCLMSRDPYCGWDQGRGIS-IYSSER-SVLQSIINPAEPHKCP--NPKDPKAPLOKYS 523
		: : : : : : : : : : : : : : : : : : : : : : : :
Db	615	YGVAGSAAFLECCQPRSPQATVKW 637
		: : : : : : : : : : : : : : : : : : : : : : : :
Qy	524	LA-PNSRYIYILSCPMESRHAITYS 545
		: : : : : : : : : : : : : : : : : : : : : : : :

Search completed: Thu Jul 8 19:31:02 1999
Job time : 51 secs.

W P E R E A

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:36:25 1999; MasPar time 28.96 Seconds
Tabular output not generated. 434.673 Million cell updates/sec

Title: >US-09-041-236-2
Description: (1-592) from US09041236.pep (37 of 45)
Perfect Score: 4411
Sequence: 1 LLLLLLWAAAASAGQHLRSGP.....AQQYGHYCFEAQEGSYFREDA 592
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.548; Variance 154.557; scale 0.236

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	661	15.0	771	13	Human semaphorin III	7.27e-51
2	620	14.1	477	13	Human collapsin.	7.26e-47
3	612	13.9	775	33	Human semaphorin.	4.36e-46
4	506	13.7	861	22	Mouse CD100 antigen.	1.67e-45
5	599	13.6	861	32	Human semaphorin W.	8.02e-45
6	544	12.3	776	32	Rat semaphorin W.	1.72e-39
7	544	12.3	862	22	Human CD100 antigen.	1.72e-39
8	463	10.5	441	13	Vaccinia virus semaphorin	1.06e-31
9	459	10.4	730	13	Grasshopper semaphorin	2.55e-31
10	428	9.7	587	32	Human semaphorin W.	2.29e-28
11	390	8.8	974	33	Human secreted protein	9.06e-25
12	378	8.6	712	13	Tribolium semaphorin	1.22e-23
13	367	8.3	650	13	Drosophila semaphorin	1.32e-22
14	362	8.2	724	13	Drosophila semaphorin	3.89e-22
15	352	8.0	930	32	Human semaphorin Y.	3.34e-21
16	325	7.4	888	25	Human semaphorin Z.	1.08e-18

17	316	7.2	887	25	W19856	Rat semaphorin Z.	7.37e-18
18	296	6.7	929	32	W57259	Rat semaphorin Y.	5.10e-16
19	157	3.6	122	13	R71385	Varicella major virus s	8.42e-04
20	107	2.4	111	32	W51315	Human semaphorin W pa	6.88e+00
21	101	2.3	2329	25	W25038	Partial BRCA2 cancer	1.87e+01
22	99	2.2	591	4	R23006	Protein transcribed f	2.60e+01
23	98	2.2	832	33	W61092	Taq DNA polymerase I	3.08e+01
24	96	2.2	832	33	W61090	Taq DNA polymerase I	4.23e+01
25	91	2.1	478	1	R04881	Recombinant elastase.	9.39e+01
26	91	2.1	598	9	R48631	Sequence of nuclear r	9.39e+01
27	92	2.1	832	33	W61091	Taq DNA polymerase I	8.02e+01
28	92	2.1	854	30	W56309	Class II S-receptor ki	8.02e+01
29	92	2.1	855	6	R29815	S receptor kinase pro	8.02e+01
30	93	2.1	4572	30	R29815	A. mediterranei rifam	6.84e+01
31	89	2.0	534	25	W25031	Partial BRCA2 cancer	1.29e+02
32	89	2.0	554	13	R66209	Novel thermostable DN	1.29e+02
33	90	2.0	638	5	R03924	E. coli HSP (dnaK).	1.10e+02
34	89	2.0	680	4	R23143	Mutant thermostable D	1.29e+02
35	89	2.0	832	13	R76693	DNA-polymerase (F73L,	1.29e+02
36	89	2.0	832	1	P90556	Purified native therm	1.29e+02
37	89	2.0	832	33	W61087	Taq DNA polymerase I	1.29e+02
38	89	2.0	832	33	W61088	Taq DNA polymerase I	1.29e+02
39	89	2.0	832	13	R76690	Taq DNA-polymerase RE	1.29e+02
40	89	2.0	832	33	W61089	Taq DNA polymerase I	1.29e+02
41	89	2.0	833	36	W59942	Amino acid sequence o	1.29e+02
42	89	2.0	833	27	W24211	Cleavage DN nuclease.	1.29e+02
43	89	2.0	833	27	W24212	Cleavage DA nuclease.	1.29e+02
44	89	2.0	3418	26	W19211	Human breast cancer s	1.29e+02
45	89	2.0	3418	26	W23287	Human breast and ovar	1.29e+02

ALIGNMENTS

RESULT 1

ID R71380 standard; Protein; 771 AA.
AC R71380;
DT 21-NOV-1995 (first entry)
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW Varicella major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PF 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 60-63; 101pp; English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87445) or varicella major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA;

Query Match 15.0%; Score 661; DB 13; Length 771;

Best Local Similarity 30.3%; Pred. No. 7.27e-51;

Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;

Db 166 rgkspsydpkiltasllldgelysgta-adfmgdrdfairtllghhhpirtedqhsrwndp 224

CC modulation and diagnosis and treatment of viral and oncological infection
CC and diseases.
SQ Sequence 650 AA;

Query Match 8.3%; Score 367; DB 13; Length 650;
Best Local Similarity 28.4%; Pred. No. 1.32e-22;
Matches 124; Conservative 104; Mismatches 158; Indels 51; Gaps 38;
Db 3 dcnqyirinvpspggrlfvcgtnsfpmcntyilsdntyleatknqgavcypdprhnt 62
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 93 DCENYITLLERSEG-LLACGTNARHPSCWN-LVNGTVVPLGEMR-GYA--PESPENSL 147
:
Db 63 svladnelysgtv-adfsgsdpiyreplqteqydals-lnapnfv-sft-qgd-f--- 114
:
QY 148 VLFEDEVYSTIRKQEVNGKIPRFRIRGESELYTSDTVMQNPQFIKATIVHQDQAYDDK 207
:
Db 115 vyffretavefincokaly-srvarvckwkggphrf-nrwtstflksrlncsppgdyp 172
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 208 IYFFREDNPD-KNPEAPLNVSRVAQLCRGQGGESSLSVSKWNTFLKAMLVCSDAATNK 266
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 173 fyfnelqsnlvegygsmssklygvfntpsnslpgsacafalqdiadtfeqgfkeq 232
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 267 -NFNRLQDYF-LLPDPGQWRTRVGVFSPNPN-Y--SVCVYSLGDI-DKV---FR-- 315
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 233 tglinsnlpvnnakvpdprrgschndsrldpdtlnfikhslmdenvpaffsqpllvrt 292
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 316 TS--S--LKGYSLSLNPAPGKCLPQPIPTTFQVADRHPEVAQRVPMGPKLTPLFH 371
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 293 stlyrtftladvaktgpktydvlfvtdhbkliksnaeasadvkvsvvleaidv 352
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 372 SK-YHQKVAHV-RMQASHGETFHVLYLTDRGTIHK-V-VEPGE-OEH--S--FA-FMI 421
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 353 ltksepirnlrvtnqydgpkdgsyddgkllivtdsqvvaqlhrchndkkitsecva 412
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 422 M-EIQPFERRAAAIQTNSLD-A-E-R-R-KLYVSSQWESVQVPLDICEYV-GGCHGCLM 473
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 413 lqdpycawdklagkors 429
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 474 SRDPYCGWDQ--GRCIS 488
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :

RESULT 14

ID R71383 standard; Protein: 724 AA.
AC R71383;
DT 21-NOV-1995 (first entry)
DE Drosophila semaphorin II protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW varicella major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Drosophila sp.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB; Q87445.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 79-82; 101pp; English.
CC The sequence of the Drosophila semaphorin II protein. The gene was
CC isolated by PCR using primers based on sequence homology between the
CC complete grasshopper and partial Tribolium and Manduca (moth) semaphorin
CC gene sequences. The products gave two different sequences, each of which
CC was used to obtain its respective complete sequence: semaphorin I
CC (Q87444) and II. The proteins encoded by the grasshopper semaphorin I
CC (Q87441), human semaphorin III (Q87442), vaccinia virus semaphorin IV,
CC Drosophila semaphorin I and II, Tribolium semaphorin I (Q87446) or
CC varicella major (smallpox) virus semaphorin IV (Q87447) genes were used to
CC generate a series of peptides (R70370-R70418), which retain semaphorin

CC receptor binding activity. The semaphorin derived or semaphorin receptor
CC derived peptides are potent modulators of nerve cell growth, immune
CC responsiveness and viral pathogenesis. They can be used in diagnosis and
CC treatment of neurological disease and neuro-regeneration, immune
CC modulation and diagnosis and treatment of viral and oncological infection
CC and diseases.
SQ Sequence 724 AA;

Query Match 8.2%; Score 362; DB 13; Length 724;
Best Local Similarity 29.0%; Pred. No. 3.88e-22;
Matches 84; Conservative 68; Mismatches 113; Indels 25; Gaps 19;
Db 265 vyffretaveyincgkavy-sriarvckdkvgknlla-hnwtvylkarlncsigsfcp 322
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 208 IYFFREDNPD-KNPEAPLNVSRVAQLCRGQGGESSLSVSKWNTFLKAMLVCSDAATNK 266
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 323 fyfneiqsgvyqlpsdkrsffat--ftstnglignsavgcfhineiaaafngkfegsssn 380
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 267 -NFNRLQDYF-LLPDPGQWRTRVGVFSPNPN-Y--SVCVYSLGDI---KV-FR---TG- 317
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 381 sawlpvlnsrvppeprgptcvndtsnlpdtvlnfshplmdkavnh-eh-nnpvvykrdl 438
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 318 S--LKGYSLSLNPAPGKCLPQPIPTTFQVADRHPEVAQRVPMGPKLTPLFHSHK-Y 374
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 439 vftklvvdkiiridlinqeyivvyvgtlnigrlykivgyvngesls-kllidifevapneal 497
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 375 HYQKVAHVHRMQAS-HGETFHVLYLTDRGTIHKVVEPGEHSEFAFNIETQPFRAAAI 433
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 498 qymeisqrkslyigtthrkdrikidlamcnrrydn-cfrc-v-rdpcygdw 544
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 434 QTMSLDAERRKLYVSSQWESVQVPLDICEV-YGGGCHGCLMSRDPYCGWD 482
||:||||: : : : ||||: : : : : : : : : : : : : : : : : :

RESULT 15

ID W57260 standard; Protein: 930 AA.
AC W57260;
DT 02-SEP-1998 (first entry)
DE Human semaphorin Y.
KW Human; semaphorin Y; nerve extension inhibitor; central nervous system;
KW peripheral nerve growth.
OS Homo sapiens.
PN W09811216-AI.
PD 19-MAR-1998.
PF 09-SEP-1997; J03167.
PR 08-AUG-1997; JP-227220.
PR 11-SEP-1996; JP-263565.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-250958/22.
DR N-PSDB; V28915, V28916.
PT DNA encoding human and rat semaphorin Y - an inhibitor of nerve
PT extension
PS Claim 1; Page 65-70; 85pp; Japanese.
CC The present sequence represents human semaphorin Y, which inhibits
CC nerve extension. Semaphorin Y genes and proteins may be used to inhibit
CC peripheral nerve growth. Semaphorin Y antagonists can be used to
CC accelerate regrowth of the central nervous system.
SQ Sequence 930 AA;
Query Match 8.0%; Score 352; DB 32; Length 930;
Best Local Similarity 28.4%; Pred. No. 3.34e-21;
Matches 132; Conservative 112; Mismatches 164; Indels 56; Gaps 44;
Db 120 ecynvirvlvpwdsqtlacqtnsfpsvcrsygitslqqegeeelsqarcpfdatqsnva 179
: : ||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 93 DCENITLLERR-SGLLACGTNARHPSCWNLVNGTVVPLGE-MRGA--PSPDENSLV 148
: : ||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 180 ifaegslsata-adfasdavyrsrlsqppl-rs-a-kydskwlrephfvgaheghd 235
: : ||||: : : : ||||: : : : : : : : : : : : : : : : : :
QY 149 LFEDEVYSTIRKQEVNGKIPR-FRRIRGESELYTSDTVMQNPQFIKATIVHQDQAYDDK 207
: : ||||: : : : ||||: : : : : : : : : : : : : : : : : :
Db 236 vyffresvedarlgkvqfsrvarvckrdmggsprald-rhwtsflklrlncs-vpgds 293
: : ||||: : : : ||||: : : : : : : : : : : : : : : : : :

Qy 208 IYFFREDNPKNPEAPLNVRVAQLCRGDOGSES-SLSVSKWNTFLKAMLVCSDAATNK 266
Db 294 tfy-f-dvlgaltgpnvnhgrsa-lfgvfttqtntspgsavcafyldieirgfggkfkeq 350
Qy 267 NFNRLOQVFL-LPDPFG-QWRDTRVYGVSFPWN-Y--SAVCVYSLGDIDKVF--R-TS- 317
Db 351 rslgawtpvsedrypsprgscagvggaalfssrdldpddvlttfikahplldpavppvt 410
Qy 318 -SLKG-Y-H-S-LPNRPCKCLP-DOQPIPTETEFQVADRHP-E-V-AQRV-EP-MGPLK 366
Db 411 hqpllttsralltqvavdgmagphsn-itvmfignsdgtvlkvltpggrsggpepille 469
Qy 367 -TPLFH--SKYHYQKVAVHRMQASHGETFHVLYLTDRGTTHKVVEPGEQEHSAFNIME 423
Db 470 eidaysparcsgktaqtarrliigleldteghrlfvafigcivylplsrcarhga-cqrs 528
Qy 424 -IQPF---R---RAA--A--IQTMSLDAERRKLYVSSQWESQVPLDLCEVYGGCH-G 470
Db 529 clasqdpvcgwhssrgcvgdirsggtdvddagngqesmehgdcqd 572
Qy 471 CLMSRDPYCGWDQGR-CISIYSE-RSVLQINPAE-PHKECPN 511

Search completed: Thu Jul 8 19:37:05 1999
Job time : 40 secs.

THIS PAGE BLANK (USPTO)

WIREFISH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:35:32 1999; Maspar time 26.98 Seconds
Tabular output not generated. 879.204 Million cell updates/sec

Title: >US-09-041-236-2
Description: (1-592) from US09041236.pep (37 of 45)
Perfect Score: 4411
Sequence: 1 LLLLLWAAAASAGHLRSGP.....AOQYGHVFCEAGEGSYFREA 592

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 49.259; Variance 86.597; scale 0.569

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1883	42.7	653	2	T03102 semaphorin homolog A3	0.00e+00
2	737	16.7	748	2	I48744 semaphorin A - mouse	6.96e-135
3	725	16.4	751	2	I48748 semaphorin E - mouse	4.12e-132
4	716	16.2	749	2	G01856 semaphorin V - human	4.92e-130
5	688	15.6	753	2	G02173 semaphorin III family	1.39e-123
6	681	15.4	772	2	A49069 collapsin - chicken	5.65e-122
7	669	15.2	772	2	I48747 semaphorin D - mouse	3.22e-119
8	667	15.1	666	2	I58169 semaphorin III - mouse	9.28e-119
9	661	15.0	771	2	D49423 semaphorin III precursor	2.21e-117
10	618	14.0	782	2	I48745 semaphorin C - mouse	1.51e-107
11	522	11.8	834	2	S66498 M-sema F protein precursor	8.34e-86
12	470	10.7	403	2	E42521 A39R protein - vaccin	3.47e-74
13	474	10.7	760	2	I48745 semaphorin B - mouse	4.48e-75
14	463	10.5	441	2	S29921 hypothetical protein	1.24e-72
15	461	10.5	730	2	JH0798 fasciclin IV precursor	3.45e-72
16	448	10.2	1074	2	JC5928 semaphorin F precursor	2.60e-69
17	389	8.8	711	2	A49423 semaphorin I precursor	2.19e-56
18	367	8.3	656	2	B49423 semaphorin I - fruit	1.25e-51
19	362	8.2	724	2	C49423 semaphorin II precursor	1.49e-50
20	293	6.6	295	2	J01775 SalL9R protein - vacc	5.64e-36
21	196	4.4	1884	2	JC4975 plexin 2 precursor -	1.09e-16
22	170	3.9	1872	2	JC4976 plexin 3 precursor	6.50e-12
23	171	3.9	1905	2	I51553 Plexin - African claw	4.30e-12

24	168	3.8	1894	2	JC4980	plexin 1 precursor -	1.48e-11
25	157	3.6	122	2	J01845	I4R protein - variola	1.27e-09
26	157	3.6	122	2	H36852	Ad3R protein - variol	1.27e-09
27	155	3.5	142	2	J01776	SalR protein - vacc	2.82e-09
28	120	2.7	775	2	E70320	polyribonucleotide nu	1.41e-03
29	105	2.4	227	2	S09922	hypothetical protein	2.13e-01
30	105	2.4	235	1	Q08EC9	HXLF4 protein - human	2.13e-01
31	102	2.3	275	2	A34866	T-cell surface protei	5.48e-01
32	103	2.3	406	2	B69064	conserved hypotherica	4.01e-01
33	99	2.2	275	2	S08464	T-cell alloantigen RT	1.38e+00
34	99	2.2	460	2	S35772	translation elongatio	1.38e+00
35	99	2.2	563	2	S78224	virulence-associated	1.38e+00
36	99	2.2	591	2	S26565	virulence-associated	1.38e+00
37	99	2.2	591	2	S09498	virulence-associated	1.38e+00
38	99	2.2	591	2	S22664	virulence-associated	1.38e+00
39	99	2.2	593	2	S15215	virulence-associated	1.38e+00
40	99	2.2	597	2	A46050	thyroid/steroid recep	1.38e+00
41	98	2.2	641	2	S32017	flagellum-associated	1.87e+00
42	98	2.2	970	2	C57282	ankyrin-related prote	1.87e+00
43	98	2.2	979	2	B57282	ankyrin-related prote	1.87e+00
44	99	2.2	1375	2	JC5148	hepatocyte growth fac	1.38e+00
45	98	2.2	1786	2	A57282	ankyrin-related prote	1.87e+00

ALIGNMENTS

RESULT 1
ENTRY #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999

ACCESSIONS T03102
REFERENCE Z14840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-653 #label ENS
#cross-references EMBL:AF005370; NID:g2337967; PID:g2337970
#length 653 #molecular_weight 73645 #checksum 5501

SUMMARY

Query Match 42.7%; Score 1883; DB 2; Length 653;
Best Local Similarity 46.3%; Pred. No. 0.00e+00;
Matches 274; Conservative 105; Mismatches 200; Indels 13; Gaps 13;

Db	57	LLMIL-SNITAAKSRFDKPRILVNLTDGFCQHRF-FGQPEPHTVLFHSLNSSDVYVCGN	114
Qy	1	LLLLLWAAAASAGHLRSGPRIFAVKGVGDVRVDFGQTEPHTVLFHEPGSSVWVGGR	60
Db	115	NTIYLFDFAHSSNASTALINTTSTHRLSSTCENFTLLHNOTDGLLAGCTNSOKPSC	174
Qy	61	GKYLDFEPPEKKNASRVTVNGSTKGCLDRDCENITLLRSEGLAGCTNARHSC	120
Db	175	W-LINNNLTQFLPKGLAPFSPSSGNLVLFDFDQNDTSTINLYKSLGS-HKFRRIAGQV	232
Qy	121	WNLVNGTVVP-LGEMRGVAPFSPDENSILVLFEGDEVSTIRK-QEYNGKIPFRIRGES	178
Db	233	ELYTSDTAMHRPQVQATAVHKNESYDDKIYFFFOENSHSDFKQPHPTVPRVQVCSDDQ	292
Qy	179	ELYTSDTVMQNPQIKATIVHQDAYDDKIYFFREDNPDKNPAPLNVSRVAQLCRGQDQ	238
Db	293	GGESLSVYKWTTLKARLACVDYDTGRIYNELQDIFWQAPENSWEETLYGLFLSPWN	352
Qy	239	GGESLSVSKWNTFLKAWLYCSDAATNKNRLOQVLLPDPSPGQWRTRVYGFSPWN	298
Db	353	FSAVCVETVKDIDHVKFTSKLKNYHKLPTPRPGCKMKNHQVPTETFOVADRYPEVADP	412
Qy	299	YSACVYVSLGIDIKVFRSSLSKGYHSSLPNRPCKLDPDQPIPTETFOVADRYPEVQR	358
Db	413	VIQNNAMFPIQSKIYITKLLVYRVEYG-GVFWATIPFLTKITTHIYVRYEDSNSTT	471

```
QY 359 VEPGPKLTPLFHSKYHYQKVAVRMOASHGETH-VLYLTDRGTIHKVYVEQGEQHSF 417
Db 472 ALNTEINPOKPAPIQNIILLNTNLKLYNSEWESEVPLDLCLSVGNCDFCFMSRDP 531
QY 418 AFNIMEIOPFRRAAIIQMSLDAERKLYVSSQWEVSOVPLDLCEVYGGCHGCLMSRDP 477
Db 532 LCTWYNNCT-S-FK-QRVSVETGTPANRTISEMCGDHYAPTIVVKHQSIPLLSNLYLSCP 588
QY 478 YCGWDOGRCLSIYSSERSVLOSINPAEPH-KE-CPNPKPKAPLQKVLAPNSRYILSCP 535
Db 589 AVSNHADYFTWKDGTETKRVHVKTKNDICILLIANSTATTATNGTHVCNMKEDS 640
QY 536 MESRHATYSWRHKNVEQSCPGHQPNCILFIENLTAQQYGHYFCEAQEGS 587

RESULT 2
ENTRY I48744 #type complete
TITLE semaphorin A - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48744
REFERENCE I48744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
#family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48744
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues 1-748 ##label RES
#cross-references EMBL:X85990; NID:g854323; PID:g854324
GENETICS semA
#gene
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 748 #molecular-weight 82894 #checksum 9017

Query Match 16.7%; Score 737; DB 2; Length 748;
Best Local Similarity 33.7%; Pred. No. 6.96e-135;
Matches 168; Conservative 112; Mismatches 166; Indels 52; Gaps 36;

Db 112 ECMFVRLHAYNTHLLACRTGAFHTCALWRWATAGGTHASGPEKLEDGKGKTPYDP 171
QY 93 DCENYITLLERRSG-LLACGTNARHESC---WNLYNGT-V-V-P--LCEMRGIAPFSP 142
Db 172 RHRPSSLVGEELYSGV-TADLMGRDFTIFRSLGQNPRLRTEPHDSRWLNPEKPVKVFWI 230
QY 143 DENSLVLFEGDEVYSTIRKQYNGK-IPRFRIRGESELYT-S-DTV-MQNPOFIKAT-I 197
Db 231 PESENPDCKIYFFRESAVEAAPAMGRMSVSRVQICRNDLGGQRL-VNKKWTFELKAR 289
QY 198 VHODAYDDKIYFFREDNPKNPE-APLNVSRVAQLCRDGGESSLSVSKWNTFLKAM 256
Db 290 LVCSVPGVEGTHFDLOQVFLLS-SR-DROTPLLYAVETSSGVFGVSVYMNVDVR 347
QY 257 LVCS--DAATNKNFNRLQDVFLLPDPSQWRDTRVGVFSPNPN-Y--SVCVYSLGDI 311
Db 348 RAFGLPLPHKEGPTHOWSYQGRVYPRPGMCPKSTEGTFSSTKDFDDVYIQGRNHPLM 407
QY 312 KVFRTS-SLK-G-----YHSSLNPNRPGKCLPDQ-QPIP-TETF--QVAD--R-HPEV 355
Db 408 YNPVLPNGG-R-PLFLQVAGYFTQTAAADRVAAADGH-YDVLFIGHDVGTVLKVISVPK 464
QY 356 AQRYEPMGPKLPLF-H--SKYHYQKVAVRMOASHGETHVLVLTDRGTIHKVVE-P- 410
Db 465 GRPNSEGLLEELQVFEDESAATISMOISSKROQLYVASRAAQAIALHRCCTALGRACAE 524
QY 411 GEQHSFAFNIMEIQPFRAAAIQTMSLDAERKLYVSSQWEVSOVPLDLCEVYGGCHG 470
Db 525 CCLARDPYCAGDWSACTRFOPTAKRRFRRODIRNGDPSTLCSDSSSVLLEKVKLVGES 584
#formal_name Homo sapiens #common_name man
#sequence_revision 06-Jun-1997 #text_change
04-Sep-1998
```

```
QY 471 CLMSRDPYCGWDOGRCLSIYSSERSVL--QsinPAEPHKECPNPKPKAPLQKVLAPNS 528
Db 585 GSFALECEPRLQAHVQW 602
QY 529 -RYILSCPMEISRATYSW 545

RESULT 3
ENTRY I48748 #type complete
TITLE semaphorin E - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48748
REFERENCE I48744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
#family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48748
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues 1-751 ##label RES
#cross-references EMBL:X85994; NID:g854331; PID:g854332
GENETICS semE
#gene
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 751 #molecular-weight 85259 #checksum 8961

Query Match 16.4%; Score 725; DB 2; Length 751;
Best Local Similarity 33.0%; Pred. No. 4.12e-132;
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;

Db 167 SFNPNTVSVMINEELFSGMYI-DFMGTDAAIFRSLTRKMQRLTDOHNSKWLSEPMFVD 225
QY 139 PFPSPDENSILVLFEGDEVYSTIRKQYNGK-IPRFRIRGESELYTS-DTV-MQNPOFIK 194
Db 226 AHVIPGTDPDNDKAVYFFFKERLTDDNRSTKQIHSIARIACPNPDTCGQRL-VNKKWTFEL 284
QY 195 ATIVHQD-QAYDDKIYFFREDNPKNPEAPLNVSRVAQLCRDGGESSLSVSKWNTFL 253
Db 285 KARLVCSVDEDDGPEHFDELDVFL-ETDNP-RTTLVYGIFTSSSVFKGSACVYHL 342
QY 254 KAMLVCS--DA-ATNKNFNRLQDVFLLPDPSQWRDTRVGVFSPNPN-Y--SVCVYSL 307
Db 343 SDIQTVFNGPFAHKEGPNHQLISYQGRIPYPRPGTCGGAFTPNMRTTKDFDDVYVTFIR 402
QY 308 GDIDKVFRTS-SLK-G-----YHSSLNPNRPGKCLPDQ-QP-I-PTETF--QVADRHP 353
Db 403 NHPLMYSISPIHRRPLIVRIGTDYKTIADVRAADG-RYHVLFLGTDRTGTQKVVV 461
QY 354 EVAQRVEPMGPL-KTPLF-H--SKYHYQKVAVRMOASHGETHVLVLTDRGTIHKVVE 409
Db 462 LPTNSSASGELILEEVEFKNHVPITTMETSSKKQOLYVSSNEGVSQVSLHRCYIGTAC 521
QY 410 -PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERKLYVSSQWEVSOVPLDLCEVYGGCG 468
Db 522 ADCCLARDPYCAGDWSCTFRPYTGKRRRRQDVRHGNPLTQCGRFNLKAYRNAAEIVQY 581
QY 469 HGCLMSRDPYCGWDOGRCLSIYSS--RSVLQSLINPAEPHKECP--NPKPDKAPLQKVL 524
Db 582 GVRNNSTIFLECAPKSPQASIKWLLQKDK 609
QY 525 A-PNSRYILSCPMEISRATYSW-RHKN 550

RESULT 4
ENTRY G01856 #type complete
TITLE semaphorin V - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
04-Sep-1998
```



```
Qy 304 VYSLGDDIKVFRTS-SLK-G--YH--S--S-LPNRPFGKCLPDQ-QPI-PTETF--QV-A 349
Db 402 FARSHPAMYNPEPINS-RPIMIKTDVYQFTQIVVDRVDAEDGQ-YDVMFIGTDIGTVL 459
Qy 350 -DR-HPEVAQRVPMGLKPTPLFHSK--YHYQKVAVHRMQASHGETFHVLYLTDRGIIH 405
Db 460 KVPSPKTEWHELEEVLEEMTVFREPTTISAMKISTKQOOLYIGSATGVSQPLHRCDV 519
Qy 406 KVEE-PGEOHSAFNAME-IOPFRAAAIOTMSLDAERRKLYVSSQWESQVPLDLCEV 463
Db 520 YGKACAECLARDPYCANDGSSCSRYFTAKRRTTRDIRNGDPLTHCSDLQHDNPSGQ 579
Qy 464 YGGCHGCLMSRDPYCGWDQRCISYSS-RSVL-QSINPAEPHKECPN-PKPDKAP-- 518
Db 580 TLEKIIYGVENSFTFLECSPKSORAIIVY-WQFQKQNDHDKVE 621
Qy 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQSC 556

RESULT 7
ENTRY 148747 #type complete
TITLE semaphorin D mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS 148747
REFERENCE #authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession 148747
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-772 #label RES
#cross-references EMBL:X85993; NID:g8543329; PID:g854330
GENETICS semD
#gene semD
#superfamily semaphorin
CLASSIFICATION #length 772 #molecular_weight 88710 #checksum 1776
SUMMARY

Query Match 15.2%; Score 669; DB 2; Length 772;
Best Local Similarity 30.2%; Pred. No. 3,22e-119;
Matches 170; Conservative 143; Mismatches 182; Indels 67; Gaps 51;

Db 58 HTFLDE-ERGRLYVGAKDHIFSNLVNIKDFQKIVWPVSX-TRDECKWAGKDLKECA 115
Qy 43 HTVLFHEPGSSVWVGGRKVLDFDFPEGKN-AS-VRTVNIQSTKGC--LDK--RDCE 95
Db 116 NFIKLEAYNOTHLYACCTGAFHPICITVIEVGHHPEDNIFKLQDSHFENGSKSPYDPKL 175
Qy 96 NYITLLRRSEG-LIACCTNARHPSCNWL-V--N--GTVVPL--G--EM-RGYAPFSPDE 144
Db 176 LTASLLDGLYSGTA-ADFMGRDFAIRFTLGDHHPHPIRTEQHDHSLNDPRFISAHLLPE 234
Qy 145 NSLVLFEGDEVYSTIRKQYNGK-IPERRIRGESELYTS--DIV-WQNPQFIKATIVHQ 200
Db 235 SDNPEDDKYVFFFRNATIGGSHGKATHARIGQICKNDFGHRSL-VNKWTFTLKARLIC 293
Qy 201 -DQAYDDKIYFFREDNPKNPEAPLVNSRVAQLCRGQDQGESLSVSKWNTFLKAMLVC 259
Db 294 SVPGPNGIDTHFDELQDVFLM-NSKDP-KNPVYGVFTTSSNIFKGSVMSDVRV 351
Qy 260 S--DA-ATNKNFNRLQDVLFPDPSGQWRDTRVYGVFSNPNW-N-Y--SAVCYVSLGDDIKV 313
Db 352 FLGPVHRDGNPNQWVPYQGRVPYPRPGTCPSKTFGGFSDTKDLPDDVITFGRSHPAMYN 411
Qy 314 FRTS-SLK-G--YH--S--S-LPNRPFGKCLPDQ-QPI-PTETF--QV-A-DR-HPEVAQ 357
Db 412 PVFPINN-RPIMIKTDVYQFTQIVVDRVDAEDGQ-YDVMFIGTDIGTVLGVSPKQETW 469
Qy 358 RVEPMGLKPTPLFHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTHKVVVE-PGQ 414
```

```
Db 470 HDLEEVLEEMTVFREPTTISAMKISTKQOOLYIGSTAGVAQLPLHRCDIYKACAECL 529
Qy 415 HSFAFNIME-IOPFRAAAIOTMSLDAERRKLYVSSQWESQVPLDLCEVYGGCHGCLM 473
Db 530 ARDPYCAWDGSSCSRYFTAKRRTTRDIRNGDPLTHCSDLQHDHGHGSPLEERIIVGV 589
Qy 474 SRDPYCGWDQRCISYSS-RSVL-QSINPAEPHKECPN-PKPDK--AP-LQ-KVSLA- 525
Db 590 ENSSTFLECSPKSORALVY-WQ 610
Qy 526 PNSRYLSC-PMESRHATYSWR 546

RESULT 8
ENTRY 158169 #type fragment
TITLE semaphorin III - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS 158169
REFERENCE #authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;
#journal Tessler-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#title Neuron (1995) 14:949-959
#title Semaphorin III can function as a selective chemorepellent to
pattern sensory projections in the spinal cord.
#cross-references MUID:95267432
#accession 158169
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-666 #label RES
#cross-references GB:L40484; NID:g703189; PID:g703190
GENETICS SemIII
#gene SemIII
#superfamily semaphorin
CLASSIFICATION #length 666 #checksum 9654
SUMMARY
```

```
Query Match 15.1%; Score 667; DB 2; Length 666;
Best Local Similarity 30.1%; Pred. No. 9,28e-119;
Matches 150; Conservative 130; Mismatches 168; Indels 50; Gaps 39;

Db 60 RGKSPYDPKLLTASLLDGLYSGTA-ANFMGRDFAIRFTLGHHPHPIRTEQHDHSLNDP 118
Qy 135 RGYAPFSPDENSLSVLFEGDEVYSTIRKQYNGK-IPERRIRGESELYTS--DIV-WQNP 190
Db 119 RFISAHLIPSDNPDQKIVYFFFRNATIGGSHGKATHARIGQICKNDFGHRSL-VNWK 177
Qy 191 QFIKATIVHQ-DQAYDDKIYFFREDNPKNPEAPLVNSRVAQLCRGQDQGESLSVSKW 249
Db 178 TFLKARLICSVCPNGIDTHFDELQDVFLM-NSKDP-KNPVYGVFTTSSNIFKGSVAVC 235
Qy 250 NTFUKAMLVCS--DA-ATNKNFNRLQDVLFPDPSGQWRDTRVYGVFSNPNW-N-Y--SAVC 303
Db 236 MYSMSDVRVRLGLGYAHRDGNPNQWVPYQGRVPYPRPGTCPSKTFGGFSDTKDLPDDVIT 295
Qy 304 VYSLGDDIKV----F--RTSS-LK--CYHSLPNPRGKCLPDQ-QPI-PTETF--QV-A 349
Db 296 FARSHPAMYNPVFPINN-RPIMIKTDVYQFTQIVVDRVDAEDGQ-YDVMFIGTDIGTVL 353
Qy 350 -DR-HPEVAQRVPMGLKPTPLFHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTH 405
Db 354 KVSVPKTEWHDLEEVLEEMTVFREPTTISAMKISTKQOOLYIGSTAGVAQLPLHRCDI 413
Qy 406 KVEE-PGEOHSAFNAME-IOPFRAAAIOTMSLDAERRKLYVSSQWESQVPLDLCEV 463
Db 414 YGKACAECLARDPYCANDGSSCSRYFTAKRRTTRDIRNGDPLTHCSDLQHDHGHG 473
Qy 464 YGGCHGCLMSRDPYCGWDQRCISYSS-RSVL-QSINPAEPHKECPN-PKPDK--AP 518
Db 474 SLEERIIVGVENSFTFLECSPKSORALVYWQFORNDRKEEIKMGDHIIRTEQGLLRS 533
Qy 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQSCPPG-HQ-SPNCILFIEN 570
```

```
Db 534 LOKKDSGNLYLCHAVEHGF 551
Qy 571 LTAQOYGHYFCEAQEGSY 588

RESULT 9
ENTRY semaphorin III precursor - human
TITLE #formal_name Homo sapiens #common_name man
ORGANISM 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
DATE 04-Sep-1998
ACCESSIONS D49423 #type complete
REFERENCE A49423 semaphorin III precursor - human
#authors Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
#journal Cell (1993) 75:1389-1399
#title The Semaphorin genes encode a family of transmembrane and
secreted growth cone guidance molecules.
#accession D49423
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-771 #label KOL
#cross-references GB:L26081; NID:g799328; PID:g436560
GENETICS
#gene GDB:SEMA1
#cross-references GDB:283448
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 771 #molecular_weight 88889 #checksum 6249

Query Match 15.0%; Score 661; DB 2; Length 771;
Best Local Similarity 30.3%; Pred. No. 2,21e-117;
Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;

Db 166 RKSPYDPKLLTASLLIDGELYSGA-ADFMGRDFAIFRTLGHHPIRTEQHDRLNDP 224
Qy 135 RGYAPFSPDENSVLFEDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNP 190
Db 225 KFSAHLISESDNPDDKVVYFFRENADIGESGKATHAGIQICKNDFGHRSL-VNWK 283
Qy 191 QFIKATIVHQ-DQAYDDKIYFFREDNPKNPEAPLVNYSVAQLCRGQGGESSLSVSKV 249
Db 284 TTEFKARLICVPGNGJDTDFEQLQVFLM-NFKDP-KNPVYGVETTSNIFKGSVAVC 341
Qy 250 NTFKAMLVCS--DA-ATNKNFNRLQDVLLPDPGQGRDTRVYGVSNPNW-Y--SVC 303
Db 342 MYSMSDVRVFLGPAHYHDPGNYQWVPYQGRVPRPGTCPSKTFGGFSDTKDLPDDVIT 401
Qy 304 VYSLGDIKDVFTS-SLK-G-YH--S--S-LPNRPGKCLPDQ-QPI-PTETF--QV-A 349
Db 402 FARSHPMYNPVPMNNRPVVIKIDVNTQFTQIVVDVDAEDGQ-YDVNFICTDVGTVLK 460
Qy 350 -DR-HPEVAQRVPMGPKLTPL-FHSKYHYQKVAVHRMQASHGETFHVLYLTDRGTIHK 406
Db 461 VYSIPKETWYLEEVLLSEMIVFERPTAISAMELSTKQOOLYIGTAGVAQLPLHRCDIY 520
Qy 407 VVE-PGQEHFAFNIME-IQFRAAAIQTMSDAERRKLYVSQWEVSQVPLDLCVEY 464
Db 521 GKAAECCLARDPYCADGSCSYFFPTAKRTRRQDTRNGDPLTHCSDLHDHNNHGHSP 580
Qy 465 GGGCHGCLMSRDPYCGWQGRGCIYSISE-RSVL-QSINPAEPHKECPNPKPDK----AP 518
Db 581 EERIYGVENSSTFLECPKSORALYVQFORNREERKEIRVDHDIIRTDQGLLRSLQ 640
Qy 519 LOKVSLA-PNSRYLSC-PMESRHATY-SW-RHKENV-EQCEPGHQ-SPNCILFIENLT 572
Db 641 QKDSGNLYLCHAVEHGFQI 658
Qy 573 AQOYGHYFCEAQEGSYFR 590

RESULT 10
ENTRY semaphorin C - mouse (fragment)
TITLE
```

```
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
ACCESSIONS I48746
REFERENCE I48746
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48746
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-782 #label RES
#cross-references EMBL:X85992; NID:g854327; PID:g854328
GENETICS
#gene semC
SUMMARY #length 782 #checksum 1571

Query Match 14.0%; Score 618; DB 2; Length 782;
Best Local Similarity 33.7%; Pred. No. 1.51e-107;
Matches 149; Conservative 96; Mismatches 148; Indels 49; Gaps 32;

Db 74 KRDCQNYIKILLPLNSSLTTCGTAAPSPLCAYIHIASTFLAQDEAGNVILEDGKHGCPF 133
Qy 91 KRDCENYIT-LLERRSEGLACGTNARHPSC-W-N-----LV-NGTV-VPLGEMRGYAPF 140
Db 134 DPNFKSTALVVDGELYTGTVS-SFQGNDPAISRSQSSRPTKTESLNWLOPAFVASATS 192
Qy 141 SPDENSVLFEDEVYSTIRKQYNGKIPRFRIRGESELYTSQTV--MQNPQFIKATIV 198
Db 193 PESLGSPIGDDKTYIFFSETGQGEFFENIT-VSRVARVCKGDEGGERVLIQ-QRWTSFL 250
Qy 199 HQDQ-A-Y--DDKIYFFREDNPD-KNPEAPLVNYSVAQLCRGQGGESSLSVSKWTFEL 253
Db 251 KAQLLCSPDGGFNFVLDVFTLNPNPQ-DWRKTLSTIGVETSQWHRGTTGSAICVFTM 309
Qy 254 KAMLVCSDAAYKNFNLQDVFLP-PDPSGQWRDTRVYGVSNPNW---Y--SAYCVYSL 307
Db 310 NDVQKAFDGLKYYNRETQWYTTHTQVTPRPGACIINSARERKINSLOLPDRVLNFL 369
Qy 308 GDIDKVF-----RUSL-LKG-Y-HSS-LPNRPGKCLPD--QQPIPTTFQVADRHPVA 356
Db 370 KDFMLDQVRSRLLLQPRARYORVAVHRVPGUHS-TYDVLFTGTGDRLUHKAVTLSR 428
Qy 357 QRVEPM-GPLKTPLF--HSKYHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVVEPGEQ 413
Db 429 VH--I-IEELQIFPOGQPVQNLNLLDSHGLLYASSHSGVQVPVANCSLYPT-CGDCLL 483
Qy 414 EHSFAFNIMEIQPFRAAAIQTMSLDAERRKLYVSQWEVSQVPLDLCLEVYGGCHGCLM 473
Db 484 ARDPYCAWGTGSACRLASLYQPD 505
Qy 474 SRDPYCGWDQRC-I-SIYSSE 493

RESULT 11
ENTRY M-sema F protein precursor - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change
DATE 10-Sep-1997
ACCESSIONS S66498
REFERENCE S66498
#authors Inagaki, S.; Furuyama, T.; Iwahashi, Y.
#journal FEBS Lett. (1995) 370:269-272
#title Identification of a member of mouse semaphorin family.
#cross-references MUID:95385809
#accession S66498
#status preliminary
#molecule_type mRNA
#residues 1-834 #label INA
#cross-references EMBL:S79463; NID:g1110598; PID:g1110599
```


Db 585 LASYHSHGRAKISEA 600
QY 540 HATYSWRH-KENVEQS 554

RESULT 14
ENTRY #type complete
TITLE hypothetical protein 15 - vaccinia virus
ORGANISM #formal_name vaccinia virus
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997

ACCESSIONS S29921
REFERENCE S29907
#authors Amegadzie, B.Y.
#submission submitted to the EMBL Data Library, January 1991
#accession S29921
##status preliminary
##molecule_type DNA
##residues 1-441 #label AME
##cross-references EMBL:X57318; NID:g62239; PID:g62254
SUMMARY #length 441 #molecular_weight 50185 #checksum 6034

Query Match 10.5%; Score 463; DB 2; Length 441;
Best Local Similarity 33.9%; Pred. No. 1.24e-72;
Matches 87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;

Db 115 LVCGTNGNPKWK-IDGSDDPKRGYAPYQNSKVYTIHSHNGC-VLSDINISK-EG-I 170
QY 109 LACGTNARHPSCNWLVNGTVVPLGEMRGYAPFSPDENSILVFEQDEVYSTIRKQYNGKI 168
Db 171 KWRREDGCGVGLTADNVIPKDG-LRGAFVDKCTYD-KVILFTDTIGSKR--I-VK 225
QY 169 PRFRRIRGE-S-ELYTSIDTMQNPQIKATIVHQDQAYDDKIYFFREDNPKNPEAPLN 226
Db 226 IPYIAQMLNDEGGSSLSHRWSTELKVELEC-DID-GRSY-R-Q-IHSTRITKD-ND 279
QY 227 VSRVAQLCRDQGGESSLSYSKNTLTKAMLYCSDAATNKNFNRLQDVLLPDPSCQWRD 286
Db 280 TILYVFFDSPYSKALCTYSMNTIKSFTSKLEGYTKOLPSPASGICLPAGKVVPHTTF 339
QY 287 TRYGVGFSPWNYSAVCVYSGLGIDKVFRTSSLKGYHSSLPNRPCKLPDQQPIPTETF 346
Db 340 EVIEKYNVLDDIKPLS 356
QY 347 QVADRHPEVAQRVEPMG 363

RESULT 15
ENTRY JH0798 #type complete
TITLE fasciclin IV precursor - American bird grasshopper
ORGANISM #formal_name Schistocerca americana #common_name American bird grasshopper
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Sep-1997

ACCESSIONS JH0798
REFERENCE JH0798
#authors Kolodkin, A.; Matthes, D.J.; O'Connor, T.P.; Patel, N.H.; Admon, A.; Bentley, D.; Goodman, C.S.
#journal Neuron (1992) 9:831-845
#title Fasciclin IV: Sequence, expression, and function during growth cone guidance in the grasshopper embryo.
#accession JH0798
##molecule_type mRNA
##residues 1-730 #label KOL
##cross-references GB:I00709; NID:g160844; PID:g160845
##experimental_source embryo
COMMENT This protein plays a role in growth cone guidance in the developing central nervous system.
KEYWORDS glycoprotein; transmembrane protein
FEATURE
1-22 #domain signal sequence #status predicted #label SIG\
23-730 #product fasciclin IV #status predicted #label MAT\
23-627 #domain extracellular #status predicted #label EXT\

628-652 #domain transmembrane #status predicted #label TMM\
653-730 #domain intracellular #status predicted #label INT\
44,71,163,267,360, #binding_site carbohydrate (Asn) (covalent) #status predicted
539

SUMMARY #length 730 #molecular_weight 81214 #checksum 5881

Query Match 10.5%; Score 461; DB 2; Length 730;
Best Local Similarity 28.0%; Pred. No. 3.45e-72;
Matches 143; Conservative 130; Mismatches 182; Indels 56; Gaps 46;

Db 32 QFGEERVORFLGNESHKDHFKLLEKDHNSLLYGARIVYNISLRDLTETEQRHSHSSC 91
QY 29 HVGQDRVD-FGOTEPHTVLFH--EPGSSVWVGGRGKVLDFEPECKN-ASVETV-NI-G 82
Db 92 AHRELCLYLGKSEDDCQNVIRVLAKIDDDRVLCGNNAYKPLCRHVALKDGDVVVEKEYE 151
QY 83 SYKSC-L-DKR--DCENYITLLERSEG-LLACGTNARHPSCWN--LYNGT-VVPLG-E 133
Db 152 GRGLCPDFPDHNSAIYSEGLYSATV-ADFGSTDP-LI-YRG-P-LRTERSDL-K-QLN 204
QY 134 MRGYAPFSPDENSILVFEQDEVYSTIRKQYNGKIPFRERRIRGESELYTSDTMQNPQFI 193
Db 205 APNFVN-TMEYNDFIFFFRETAIVEYINGCKAIY-SRVARVCKHDKGGPHQFG-DRWTSF 261
QY 194 KATIVHQDQAYDDKIYFFREDNPD-KNPEAPLNVSRAQLCRDQGGESSLSVSKWNTF 252
Db 262 LKSLRMCSPGQDYPPFYFNEIQSTSDIEGNYGGO-VEKLIYGVFTTPVNSIGGSAYCAFS 320
QY 253 LKAMLYCSDAATNK-NFNRLQDV-FLLPDP-SGQWRDTRYGVFSPNPN-Y--SAVCVYS 306
Db 321 MKSILESFGDPFKEQETMNSNMLAVPSLKVPPEPRQCVNDSRTLPDVSVNFVKSHTLMD 380
QY 307 L-G-D-IDKVERT-SSLKG-YHS--SL-PNPRCKCLPDQOPIPTETFQVADRHEVA 356
Db 381 EAVPAFFT-RPILIRISLOYRFTKIADVQOVRTPKGKAYDVLFIGTDDGKVIKALNSASF 439
QY 357 QRVEPMGLPKTLFH-S-KYHYQKVAV-HRMQASHGETFHVLYLTTRDTGTHKRVPEPG-- 411
Db 440 DSSDVTVDVVIEELQVLPPGPVKNLVYVMRMDGDDSKLVVVSDETLAKLHRCGSDKIT 499
QY 412 EQEHSFAFNIME-IQPFRAAAIQTM--SLDAERKRLIVSSQWEVVSQVPLDLCVEYG-G 466
Db 500 NCRECVSLQDPYCAMDNVELKCTAVGSPDWS 530
QY 467 GCHGCLMSRDPYCGWDQ-G-RCISIIYSERS 495

Search completed: Thu Jul 8 19:36:07 1999
Job time : 35 secs.

THIS PAGE BLANK (USPTO)


```
::
188 ON 189

RESULT 8
ID EF1A_TRIRE STANDARD; PRT; 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEF1.
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYrenomYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-QM9414 / RUTC-30;
RX MEDLINE; 94124021.
RA NAKARI T., ALATALO E., PENTTILA M.;
RT "Isolation of Trichoderma reesei genes highly expressed on glucose-
RT containing media: characterization of the tefl gene encoding
RT translation elongation factor 1 alpha.";
RL GENE 136:313-318(1993).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; 223012; G312887; -.
DR PIR; S35772; S35772.
DR PROSITE; PS00301; EFACITOR_GTP; 1.
DR PFAM; PF00009; GTP_EFTU; 1.
DR HSSP; P07157; LAIP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 49830 MW; B9ABAB2 CRC32;

Query Match 2.2%; Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 3.32e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;

Db 210 YKGWEKETKACFTG-KTLLLEAIDSIIEPKR-PTDKPLRLPQDV 252
| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 478 YGWDQGRCSIIYSERSVLOSINPAEPHKCPNPKAPLQKV 522
| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
ID VL2_HPV5B STANDARD; PRT; 518 AA.
AC P25540;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MINOR CAPSID PROTEIN L2.
GN L2.
OS HUMAN PAPILLOMAVIRUS TYPE 5B.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91306467.
RA YABE Y., SAKAI A., HITSUMOTO T., KATO H., OGURA H.;

RT "A subtype of human papillomavirus 5 (HPV-5b) and its subgenomic
RT segment amplified in a carcinoma: nucleotide sequences and genomic
RT organizations.";
RL VIROLOGY 183:793-798(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; D90252; D1015003; -.
DR PIR; H40480; P2WLB5.
DR PFAM; PF00513; late_protein_L2; 1.
KW COAT PROTEIN; LATE PROTEIN.
SQ SEQUENCE 518 AA; 56836 MW; D0ADC243 CRC32;

Query Match 2.2%; Score 95; DB 1; Length 518;
Best Local Similarity 27.1%; Pred. No. 1.28e+00;
Matches 16; Conservative 17; Mismatches 25; Indels 1; Gaps 1;

Db 219 GDITDIIIEIPSRYTEIEEPTPRRSSTPLRPNQSVGRRGFSLTNRRLVQOVVD 277
| | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 402 GTIHKVVEPGGEHSFAFNIEMIQFRRAAAIQTMSLDAERKLY-VSSQWEVSQVPLD 459
| | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 10
ID VL2_HPV05 STANDARD; PRT; 518 AA.
AC P05918;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MINOR CAPSID PROTEIN L2.
GN L2.
OS HUMAN PAPILLOMAVIRUS TYPE 5.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87207670.
RA ZACHOW K.R., OSTROW R.S., FARAS A.J.;
RT "Nucleotide sequence and genome organization of human papillomavirus
RT type 5.";
RL VIROLOGY 158:251-254(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; M17463; G484221; -.
DR PIR; B26277; P2WLB5.
DR PFAM; PF00513; late_protein_L2; 1.
KW COAT PROTEIN; LATE PROTEIN.
SQ SEQUENCE 518 AA; 56704 MW; B7199004 CRC32;

Query Match 2.2%; Score 95; DB 1; Length 518;
Best Local Similarity 27.1%; Pred. No. 1.28e+00;
Matches 16; Conservative 17; Mismatches 25; Indels 1; Gaps 1;

Db 219 GDITDIIIEIPSRYTEIEEPTPRRSSTPLRPNQSVGRRGFSLTNRRLVQOVVD 277
| | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 402 GTIHKVVEPGGEHSFAFNIEMIQFRRAAAIQTMSLDAERKLY-VSSQWEVSQVPLD 459
| | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 11
ID VRP2_SALTY STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
```

"Nucleotide sequences of genes encoding 32 kDa and 70 kDa

```
Query Match      2.2%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.32e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGLHLLGKTAARLSDPQAASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RDCENYITLLRRSEGLACGTNARHPSCWNLVNGTVPLGEMRGYAPFSPDENSILVLF 151
Db 220 GNEAGRDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPFRIRGESELY 181

RESULT 14
ID VRP2_SALDU STANDARD; PRT; 593 AA.
AC P24419;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN VSDC.
OS SALMONELLA DUBLIN.
OC PLASMID PSDL2.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
[1]
RX MEDLINE; 91251759.
RA KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINEY D.;
RT "Molecular analysis of the virulence locus of the Salmonella dublin
plasmid pSDL2.";
RL MOL. MICROBIOL. 5:307-316(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X56727; G47839;
CC PIR; S15215; S15215.
CC KW PLASMID; VIRULENCE.
CC FT DOMAIN 367 375 POLY-PRO.
CC SEQUENCE 593 AA; 65613 MW; ATCE57B5 CRC32;

Query Match      2.2%; Score 99; DB 1; Length 593;
Best Local Similarity 23.3%; Pred. No. 3.32e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGLHLLGKTAARLSDPQAASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RDCENYITLLRRSEGLACGTNARHPSCWNLVNGTVPLGEMRGYAPFSPDENSILVLF 151
Db 220 GNEAGRDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPFRIRGESELY 181

RESULT 15
ID AMOH_ARTGO STANDARD; PRT; 683 AA.
AC Q59118;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HISTAMINE OXIDASE (EC 1.4.3.6) (COPPER AMINE OXIDASE).
OS ARTHROBACTER GLOBIFORMIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIADAE;

ACTINOMYCETALES; MICROCOCCEAE; MICROCOCCACEAE; ARTHROBACTER.
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-IFO 12137 / ATCC 8010;
RX MEDLINE; 95181469.
RA CHOI Y.-H., MATSUZAKI R., FUKUI T., SHIMIZU E., YORIFUJI T., SATO H.,
OZAKI Y., TANIZAWA K.;
RT "Copper/topa quinone-containing histamine oxidase from Arthrobacter
globiformis. Molecular cloning and sequencing, overproduction of
precursor enzyme, and generation of topa quinone cofactor.";
RL J. BIOL. CHEM. 270:4712-4720(1995).
CC -!- FUNCTION: OXIDIZE HISTAMINE. OTHER AMINES INCLUDING
PHENETHYLAMINE, TYRAMINE, TRYPTAMINE, PUTRESCINE, AND BENZYLAMINE
ALSO SERVE AS SUBSTRATE.
CC -!- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -
RCHO + NH(3) + H(2)O(2).
CC -!- COFACTOR: COPPER AND TOPAQUNINONE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- INDUCTION: BY HISTAMINE.
CC -!- PTM: THE TOPA QUINONE IS GENERATED BY COPPER-DEPENDENT
AUTOXIDATION OF A SPECIFIC TYROSYL RESIDUE.
CC -!- SIMILARITY: BELONGS TO THE COPPER/TOPAQUNONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D38508; G594747;
CC PROSITE; PS01164; COPPER_AMINE_OXID_1; 1.
CC PROSITE; PS01163; COPPER_AMINE_OXID_2; FALSE_NEG.
CC PFAM; PF01179; Cu_amine_oxid; 1.
CC HSSP; P46881; 1AVL.
CC KW OXIDOREDUCTASE; COPPER; TPQ.
CC FT INIT_MET 0
CC BINDING 401 401 TOPAQUNONE.
CC METAL 450 450 COPPER (POTENTIAL).
CC METAL 452 452 COPPER (POTENTIAL).
CC METAL 611 611 COPPER (POTENTIAL).
CC SEQUENCE 683 AA; 74979 MW; 82C8CB73 CRC32;

Query Match      2.2%; Score 95; DB 1; Length 683;
Best Local Similarity 28.2%; Pred. No. 1.28e+00;
Matches 31; Conservative 26; Mismatches 44; Indels 9; Gaps 9;

Db 494 TLLARESEAVRD-ADGAK-GRVWHISNPDLSNLHGLHPVGYTLY-PEGNP-TLAMADDS-S 548
QY 99 TLLERRSEGLACGTNARHPSCWNLVN-GTVVPLGEMRGYAPFSPDENSILVLFEGDEVYS 157
Db 549 IASRAAFARHHLWYTR-HAEELYAAGDFVNOHPGGAFLPAYVAQDRDID 597
QY 158 TIRKQYNGKIPRRIRGESELYTS-DTVMNQFQ-FIKATIVHQDQAYD 205

Search completed: Thu Jul 8 19:34:05 1999
Job time : 33 secs.
```

THIS PAGE BLANK (USPTO)

[M][A][P][S][R][C][H] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 19:34:24 1999; Maspar time 39.05 Seconds
Tabular output not generated.
827.374 Million cell updates/sec
Title: >US-09-041-236-2
Description: (1-592) from US09041236.pep (37 of 45)
Perfect Score: 4411
Sequence: 1 LLLLLWAAAASAGHLRSGP.....AQQYGHVCFEAQEGSYFREA 592

Scoring table: PAM 150
Gap 11
Searched: 179066 seqs, 54579741 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: sptrembl9
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 48.819; Variance 74.839; scale 0.652
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	4411	100.0	666	4	SEMAPHORIN L.	0.00e+00
2	2398	54.4	393	11	SEMAPHORIN L (FRAGMENT)	0.00e+00
3	1883	42.7	653	14	SIMILAR TO GENBANK ACC	0.00e+00
4	737	16.7	748	11	SEMAPHORIN A PRECURSOR	3.40e-155
5	725	16.4	751	11	SEMAPHORIN E PRECURSOR	5.15e-152
6	716	16.2	749	4	SEMAPHORIN V.	1.24e-149
7	716	16.2	750	4	SEMAPHORIN V.	1.24e-149
8	710	16.1	751	4	SEMAPHORIN E.	4.81e-148
9	705	16.0	751	13	COLLAPSID 3.	1.01e-146
10	688	15.6	753	4	SEMAPHORIN III FAMILY	3.12e-142
11	682	15.5	754	11	SEMAPHORIN IV ISOFORM	1.19e-140
12	682	15.5	785	11	SEMAPHORIN IV ISOFORM	1.19e-140
13	681	15.4	772	13	COLLAPSID.	2.19e-140
14	681	15.4	785	4	SEMAPHORIN IV.	2.19e-140
15	679	15.4	785	4	SEMAPHORIN D.	7.37e-140
16	669	15.2	772	11	SEMAPHORIN D PRECURSOR	3.18e-137
17	668	15.1	772	11	SEMAPHORIN III/COLLAPS	5.83e-137
18	661	15.0	771	4	SEMAPHORIN-HII.	4.05e-135
19	661	14.0	775	11	SEMAPHORIN H.	7.78e-124
20	618	14.0	782	11	SEMAPHORIN C (SEM C) (7.78e-124

SEMAPHORIN J (SEMAPHOR K1AA0331.
COLLAPSID-2.
COLLAPSID-3 (FRAGMENT)
SEMAPHORIN.
SEMAPHORIN I (M-SEMA F
COLLAPSID-5 (FRAGMENT)
COLLAPSID-4 (FRAGMENT)
SEMAPHORIN B PRECURSOR
FASCTICLIN IV.
SEMAPHORIN F HOMOLOG.
SEMAPHORIN F PRECURSOR
SEMAPHORIN G PRECURSOR
SEMAPHORIN VIA.
SEMAPHORIN F (FRAGMENT
SEMAPHORIN-I PRECURSOR
CESENA.
SEMAPHORIN-I.
SEMAPHORIN-II.
SEMAPHORIN IV HOMOLOG
SEMAPHORIN N (SEMAPHOR
SEMAPHORIN Z.
SIMILAR TO SEMAPHORIN-
K1AA0463 PROTEIN (FRAG

SEMAPHORIN J (SEMAPHOR K1AA0331.
COLLAPSID-2.
COLLAPSID-3 (FRAGMENT)
SEMAPHORIN.
SEMAPHORIN I (M-SEMA F
COLLAPSID-5 (FRAGMENT)
COLLAPSID-4 (FRAGMENT)
SEMAPHORIN B PRECURSOR
FASCTICLIN IV.
SEMAPHORIN F HOMOLOG.
SEMAPHORIN F PRECURSOR
SEMAPHORIN G PRECURSOR
SEMAPHORIN VIA.
SEMAPHORIN F (FRAGMENT
SEMAPHORIN-I PRECURSOR
CESENA.
SEMAPHORIN-I.
SEMAPHORIN-II.
SEMAPHORIN IV HOMOLOG
SEMAPHORIN N (SEMAPHOR
SEMAPHORIN Z.
SIMILAR TO SEMAPHORIN-
K1AA0463 PROTEIN (FRAG

ALIGNMENTS

RESULT	ID	PRELIMINARY;	PRT;	666 AA.
1	075326			
AC	075326;			
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L.			
GN	SEMAL.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC	CATARRHINI; HOMINIDAE; HOMO.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 98389619.			
RA	LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;			
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA			
RT	viruses."			
RL	GENOMICS 51:340-350(1998).			
DR	EMBL; AF030698; G3523115; .			
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;			

Query Match 100.0%; Score 4411; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	33	LLLLWAAAASAGHLRSGRIFAVKMGHVGDVDFGQTEPHVLFHEPGSSVWVGR	92
QY	1	LLLLWAAAASAGHLRSGRIFAVKMGHVGDVDFGQTEPHVLFHEPGSSVWVGR	60
Db	93	GKYLDFEGRNASVRYTNIKSGCLDKRDCENYITLLERSEGLLACGTNARHPS	152
QY	61	GKYLDFEGRNASVRYTNIKSGCLDKRDCENYITLLERSEGLLACGTNARHPS	120
Db	153	WNLVNGTVVPLGEMRGYAPFSPDENSELVLFEGDEVYSTIRKOEYNGKIPFRIRGESEL	212
QY	121	WNLVNGTVVPLGEMRGYAPFSPDENSELVLFEGDEVYSTIRKOEYNGKIPFRIRGESEL	180
Db	213	YTSDTVMQNPQFKATIVHQDQAYDDKIYFFREDNPKNPAPLNVSVAQLCRGDQGG	272
QY	181	YTSDTVMQNPQFKATIVHQDQAYDDKIYFFREDNPKNPAPLNVSVAQLCRGDQGG	240
Db	273	ESSLSVSKNTFLKMLVCSDAATNKNRNLQDVFLLPDPGQWRTRVYGFSPNPNYS	332
QY	241	ESSLSVSKNTFLKMLVCSDAATNKNRNLQDVFLLPDPGQWRTRVYGFSPNPNYS	300

```
Db 333 AVCVYSLGDIKVFRTSSLGKGYHSSLPNRPNGKCLPDQOPIPTETFOVADRHPVEAQRVE 392
QY 301 AVCVYSLGDIKVFRTSSLGKGYHSSLPNRPNGKCLPDQOPIPTETFOVADRHPVEAQRVE 360
Db 393 PMGPLKTLPLFHSKYHYQKVAVHRMQASHGETHFHLYLITDRTGTHIKVVEPGEQESHFAFN 452
QY 361 PMGPLKTLPLFHSKYHYQKVAVHRMQASHGETHFHLYLITDRTGTHIKVVEPGEQESHFAFN 420
Db 453 IMEIOPFRRAAAIQTMSLDAERKRLVYSSQWVSDVPLDCLCEVGGCHGCLMSRDPYCG 512
QY 421 IMEIOPFRRAAAIQTMSLDAERKRLVYSSQWVSDVPLDCLCEVGGCHGCLMSRDPYCG 480
Db 513 WQOGRCSIYSSERSVLOSINPAEPHKECPNPKAPLQKVSLAPNSRYILSCPMESRH 572
QY 481 WQOGRCSIYSSERSVLOSINPAEPHKECPNPKAPLQKVSLAPNSRYILSCPMESRH 540
Db 573 ATYSRHRKENVQSCPEGHOSPCILFTENLTAQOYGHYFCEAQEGSYFREA 624
QY 541 ATYSRHRKENVQSCPEGHOSPCILFTENLTAQOYGHYFCEAQEGSYFREA 592
RESULT 2
ID 088371 PRELIMINARY: PRT: 393 AA.
AC 088371:
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN L (FRAGMENT).
GN SENAL.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSEER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030699; G3523117; -.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;
Query Match 54.4%; Score 2398; DB 11; Length 393;
Best Local Similarity 90.4%; Pred. No. 0.00e+00;
Matches 329; Conservative 16; Mismatches 15; Indels 4; Gaps 2;
Db 33 LLLVFWAAASQAQHSRSGPRISAVWKG---QDHVDFSQPEPHTVLFHEPGSFSVWVGG 89
QY 1 LLLLLWAAASQAQHLRSGPRIFAVWKGHVQDRVDFGQTEPHTVLFHEPGSSVWVGG 60
Db 90 GKVIYHFNPEGKNASRVNTGSTGSKQDQDCQNYITLLRRNGLLVCGTNARKPSC 149
QY 61 GKVIYDFPEGKNASRVNTGSTGSKLDRDCENYITLLRRSEGLACGTNARHPSC 120
Db 150 WNLVNDVYVMSLGMKGAPSPDENSILVLFEGDEVYSTIRKQYNGKIPFRERRIGESE 209
QY 121 WNLVNGTVV-PLGEMRGYAPSPDENSILVLFEGDEVYSTIRKQYNGKIPFRERRIGESE 179
Db 210 LYTSVTMGNQPIKATIVHQDAYDDKIYFFREDNPKNPEAPLNSRYAQLCRGDQ 269
QY 180 LYTSVTMGNQPIKATIVHQDAYDDKIYFFREDNPKNPEAPLNSRYAQLCRGDQ 239
Db 270 GESSLSVSKWNTFLKAMLVCSDAATNRNRLQDVLFLPDSPGQWRDTRVYGVFSNPWNY 329
QY 240 GESSLSVSKWNTFLKAMLVCSDAATNRNRLQDVLFLPDSPGQWRDTRVYGVFSNPWNY 299
Db 330 SAVCVYSLGDIKVFRTSSLGKGYHSSLPNRPNGKCLPDQOPIPTETFOVADRHPVEAQRV 389
QY 300 SAVCVYSLGDIKVFRTSSLGKGYHSSLPNRPNGKCLPDQOPIPTETFOVADRHPVEAQRV 359
Db 390 EPMG 393
```

```
QY 360 EPMG 363
||||
RESULT 3
ID 064906 PRELIMINARY: PRT: 653 AA.
AC 064906:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALCELAHPHINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C500;
RX MEDLINE; 97201573.
RA ENSSEER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C500;
RX MEDLINE; 97404659.
RA ENSSEER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=C500;
RA ENSSEER A., PFLANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; U18243; G1000717; -.
DR EMBL; AF005370; G2337970; -.
SQ SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;
Query Match 42.7%; Score 1883; DB 14; Length 653;
Best Local Similarity 46.3%; Pred. No. 0.00e+00;
Matches 274; Conservative 105; Mismatches 200; Indels 13; Gaps 13;
Db 57 LLMIL-SAITAARSFIDKPLIVNLTDGFGQHRF-FGQEPHTVLFHLSNDSVYVGGN 114
QY 1 LLLLLWAAASQAQHLRSGPRIFAVWKGHVQDRVDFGQTEPHTVLFHEPGSSVWVGG 60
Db 115 NTIYLFDAHSSNASTALINITSTHNRSLSTCENITILLHNOTDGLACGTNSQKPS 174
QY 61 GKVIYDFPEGKNASRVNTGSTGSKLDRDCENYITLLRRSEGLACGTNARHPSC 120
Db 175 W-LNNLTTQFLGPKGLAPFSPSSGNLVLFDONDYTYNLYKSLGS-HKFRRIAGOV 232
QY 121 WNLVNGTVV-PLGEMRGYAPSPDENSILVLFEGDEVYSTIRK-QEYNGKIPFRERRIGES 178
Db 233 ELYTSDTAMHRPQVQATAVHKNESYDDKIYFFQENSHSDFKQFPHTPRVGVQCSDDQ 292
QY 179 ELYTSDTVMGNQPIKATIVHQDAYDDKIYFFREDNPKNPEAPLNSRYAQLCRGDQ 238
Db 293 GGESSLSVYKWTFLKARLACVDYDTRGRYNELODIWQAPENSWEETLYIGLFLSPWN 352
QY 239 GGESSLSVSKWNTFLKAMLVCSDAATNRNRLQDVLFLPDSPGQWRDTRVYGVFSNPWN 298
Db 353 FSACVFTVXDIDHVFKTSKLNKYNHKLPTPRPGCKMKNHQPVTETFOVADRHPVEADP 412
QY 299 YSACVYSLGDIKVFRTSSLGKGYHSSLPNRPNGKCLPDQOPIPTETFOVADRHPVEAQR 358
Db 413 VYQKNNAMFPIQSKYITKLYLVVEYIG-GVFWATIFYLTITIKGTHIHYVYEDSNST 471
QY 359 VEPMGPLKTLPLFHSKYHYQKVAVHRMQASHGETFH-VLYLTTDRGTTHIKVVEPGEQESH 417
Db 472 ALNLTENPQKPAPIQNILLDNTNKLKLYNSEWESEVPLDCLSVYGNDCFCFSMRDP 531
QY 418 AFNIMEIOPFRRAAAIQTMSLDAERKRLVYSSQWVSDVPLDCLCEVGGCHGCLMSRDP 477
```

Db 532 LCTWYNTC-S-PK-QRVSVETGCPANRTISEMCGDHYAVTYVYKHQVSIPLLSNYSILSCP 588
QY 478 YCGDQGRCSISYSSRSVLOSINPAEPH-KE-CPNPKPKAPLOKQVSLAPNSRYIYLSCP 535
Db 589 AVSNHAYFWTKDFTKRCVKHVKHNDICILLIANSTATTATNGTVCNKKEDS 640
QY 536 MESRHATYSNRHENVEQSCPEGHQSPNCILFIENLTAAQQYGHVFCBAQEGS 587

RESULT 4
ID Q62177 PRELIMINARY; PRT; 748 AA.
AC Q62177;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN A PRECURSOR (SEM A).
GN SEMA OR SEMA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC -1- SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC -1- LOW LEVELS FOUND BETWEEN DAYS 10-12.
CC EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL
CC BIRTH.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85990; G854324; -.
DR MGD; MGI:107561; SEMA.
DR PFAM; PF00047; Ig; 1.
KW SIGNAL: IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KW DEVELOPMENTAL PROTEIN.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 748 SEMAPHORIN A.
FT DOMAIN 586 649 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 748 AA; 82894 MW; A7E53A8D CRC32;

Query Match 16.7%; Score 737; DB 11; Length 748;
Best Local Similarity 33.7%; Pred. No. 3.40e-155;
Matches 168; Conservative 112; Mismatches 166; Indels 52; Gaps 36;

Db 112 ECMNFVLLHAYNTHLLACRTAGFHTCALWRWATAGGTHASTGPEKLEDGKGKTPYDP 171
QY 93 DCENYITLLERSRG-LLACGTNARHSC-----WNLVNGT-V-P-LGEMRGYAFSP 142
Db 172 RHRPSPVLLGEELYSV-TADMGDRFTIFRSLGQNSPLRTEPHDSRWLNKPEKVFYWI 230
QY 143 DENSLVLFEGDEVYSTIRKQYNGK-IPRFRIRGESELYT-S-DTV-MQNPFQIKAT-I 197
Db 231 PESENPDCKIYFFRESAEVAEPAMGRMSVSRVQICRNDLGGQSL-VNKWTFELKAR 289
QY 198 VHODQAYDDKIYFFREDNPNKPE-APLNVSRVAQLCRDGGOGGESSLSVSKNTEFLKAM 256
Db 290 LVCSPVGEVGDTHDQLQDVFLLS-SR-DROTPLLYAVESTSSGVFGSACVYVSMNDVR 347
QY 257 LVCS--DAATNKNRNLQDVFLLPDPGQMRDTRVYGVSNPNW-Y--SACVYSLGDID 311
Db 348 RAFLGLPLHKEGPTHQWYVQGRVYPRGPCSKTEGTSSTKDFPDVDFQFGRNPLM 407
QY 312 KVFRTS-SLK-G-----YHSLPLNRPCKLPDQ-QPIP-TETF--QVAD--R-HPEV 355
Db 408 YNPVPLMGG-R-PLFLOVAGAGYTFQTAAADRVAAADGH-YDVLFIGDVTGLKVISVPK 464

QY 356 AQRVEPMGLKTPLF-H--SKYHYQKVAVHVRMOASHGETHVLVLTDRGTIHKVVE-P- 410
Db 465 GRRPNSGLLEELQVDFEDSAATISMQISSKROOLYVASRAAQAIALHRCRTALGRACAE 524
QY 411 GEQHSFAFNIMEIQPFRAAAQTMSLDAERRKLYVSSQWEYSQVPLDLCEVYGGCGHG 470
Db 525 CCLARDPYCWDGSACTRFQPTAKRRFRRODIRNGDSTLCSDSSHSVLEKVLGVES 584
QY 471 CLMSRDPYCGWDGRCISYSSERSVL--QSINPAEPHKECPNPKDKAPLQKVS LAPNS 528
Db 585 GSALFECEPESLQAHVQW 602
QY 529 -RYLSCPMSRHTATYSW 545

RESULT 5
ID Q62181 PRELIMINARY; PRT; 751 AA.
AC Q62181;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E PRECURSOR (SEM E).
GN SEMAE OR SEME.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRI;
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC -1- SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC -1- MAXIMUM EXPRESSION BETWEEN DAYS 10-12 WITH MODERATE LEVELS FROM
CC DAY 13 UNTIL BIRTH.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85994; G854332; -.
DR MGD; MGI:107557; SEMAE.
DR PFAM; PF00047; Ig; 1.
KW SIGNAL: IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KW DEVELOPMENTAL PROTEIN.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 751 SEMAPHORIN E.
FT DOMAIN 587 649 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 751 AA; 85259 MW; B28D6CFE CRC32;

Query Match 16.4%; Score 725; DB 11; Length 751;
Best Local Similarity 33.0%; Pred. No. 5.19e-152;
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;

Db 167 SFNPVNTVSMINEELFSGMYI-DEMGTDAAIFRSUTKRMQLRTDQHNKSLSEPMFVD 225
QY 139 PFPDENSLVLFEGDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DIV-MQNPFQIK 194
Db 226 AHVPGCTDPNDARVYFFFKERLRTDNNRSTKQIHSMIARICPNDTGGQSL-VNKWTFEL 284
QY 195 ATIVHOD-QAYDKIYFFREDNPNKPEAPLNVSRVAQLCRDGGOGGESSLSVSKNTEFL 253
Db 285 KARLVCSVTDEDPETHFDELEDVFL-ETDNP-RTLVYIGITSSSVFKGSACVYVHL 342
QY 254 KAMLVCS--DA-ATNKNRNLQDVFLLPDPGQMRDTRVYGVSNPNW-Y--SACVYSL 307
Db 343 SDIOTVFNPGFAKEGPNHQLISVQGRIPYPRCTCPGGAFTPNMRTTKDFPDVVFIR 402
QY 308 GDIDKVFRTS-SLK-G-----YHSLPLNRPCKLPDQ-QP-I-PTETF--QVADRHP 353
Db 403 NHPLMYSISPIHRRRLIVRIGTDYKTKIAVRVNAADG-RYHVLFLGTRDGTQVQVWV 461


```
RESULT 8
ID Q99985; PRELIMINARY: PRT: 751 AA.
AC Q99985;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA YAMADA T., ENDO R., GOTOH M., HIROHASHI S.;
RT "Identification of semaphorin E as a non-MDR drug resistance gene of
human cancers.";
RL PROC. NATL. ACAD. SCI. U.S.A. 94:14716-14718(1997).
DR EMBL: AB000220; D1033360; -.
SQ SEQUENCE 751 AA; 85207 MW; F89FB934 CRC32;

Query Match 16.1%; Score 710; DB 4; Length 751;
Best Local Similarity 33.6%; Pred. No. 4,81e-146;
Matches 151; Conservative 104; Mismatches 151; Indels 43; Gaps 30;

Db 167 SFPNNTVNTVMINEELFSGMYI-DFMGTDAAIFSLTKRNAVRTDQHNKWLSEPMFVD 225
Qy 139 PFPSPDLSLVFEGDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNPFQFIK 194
Db 226 AHVDPGTDNDKAVYFFKFKELTNNSTKQIHSMIARICPNDTGGRLSL-VNKWTFIL 284
Qy 195 ATIVHQD-QAYDDKIIYFFREDNPKNEAPLNVSRVAQLCRGDOGGESSLSVKWNTFL 253
Db 285 KARLVCSVTDEDPETHFEDEVELL-ETDNP-RTTLVYGIFTSSSVKFGSAVCVYHL 342
Qy 254 KAMLVCS--DA-ATKNFNRLQDVELLPDPSCQWRDTRVYGVFSPWN-Y--SAVCVYSL 307
Db 343 SDIQTVFNGPFAHKEGPNHQLISYGRIPYPRGTCPGAFTPNMRTTKEFPDDVVTFIR 402
Qy 308 GDIDKVFRTS-SLK-G-----YHSSLPNPRPGKLPDQ-QP-IPET-F--QVADRH 351
Db 403 NHPLMYNPIYTHKRPLIIRIGTDYKTKIAVDVRYNAADG-RYHVLFLGTDGRGVOKVV 460
Qy 352 -HPEVAQRVEPN-G-PLKTPFHSYKHQKVAHRMQASHGETFHVLYLTDRGTHIKVV 408
Db 461 VLPTNNSVSGELILEEVLKHPHAPITMKISSKKQQLYVSSNEGVSQVSLRHCHYIGTA 520
Qy 409 E-PGEOEHSAFNIMEIOFRAAAIQTMSLDAERRKLYVSSQWESVQVPLDCEVYGG 467
Db 521 CADCCCLARDPYCAWDGNSCRYPYTKRRSRQDVRHGNPLTCRCGFNLKAYRNAAEIVQ 580
Qy 468 HGCLMSRDPYCGWDOGRCSISYSE--RSVLQSNPAEPHKECP--NPKPKAPLQKVS 523
Db 581 YGVKNNTTFLECAPKSPQASIKWLQKDK 609
Qy 524 LAP-NSRYLSCPMESRHATYSNR-HKEN 550

RESULT 9
ID O42236; PRELIMINARY: PRT: 751 AA.
AC O42236;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COLLAPSEN 3.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASTANIDAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA FEINER L., KOPPEL A.M., KOBAYASHI H., RAPER J.A.;
RL NEURON 19:0-0(1997).
DR EMBL: AF022946; G2522204; -.
SQ SEQUENCE 751 AA; 85432 MW; 0B9BA2F7 CRC32;

Query Match 15.6%; Score 688; DB 4; Length 753;
Best Local Similarity 31.4%; Pred. No. 3,12e-142;
Matches 139; Conservative 121; Mismatches 140; Indels 42; Gaps 30;

Db 171 PYDPKLDTASALINEELVAGYVI-DFMGTDAAIFTLKQTMRTDQVNSRWLNDFSIH 229
Qy 139 PFPSPDLSLVFEGDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNPFQFIK 194
Db 230 RELIPDPSAENDDKLYFFERSAE-APOSAPVAYRIGRLNDDGHCCL-VNKNSTFLK 287
Qy 195 ATIVHQDQAYDDKIIYFFREDNPKNEAPLNVSRVAQLCRGDOGGESSLSVKWNTFLK 254

Query Match 16.0%; Score 705; DB 13; Length 751;
Best Local Similarity 33.5%; Pred. No. 1,01e-146;
Matches 150; Conservative 107; Mismatches 150; Indels 41; Gaps 28;

Db 167 SFPNNTVNTVMINEELFSGMYI-DFMGTDAAIFSLTKRNAVRTDQHNKWLSEPIFVD 225
Qy 139 PFPSPDLSLVFEGDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNPFQFIK 194
Db 226 AHVDPGTDNDKAVYFFKFKELTNNSTKQIHSMIARICPNDTGGORSL-VNKWTFIL 284
Qy 195 ATIVHQD-QAYDDKIIYFFREDNPKNEAPLNVSRVAQLCRGDOGGESSLSVKWNTFL 253
Db 285 KARLVCSVTDEDPETHFEDEVELL-ETDNP-RTTLVYGIFTSSSVKFGSAVCVYHL 342
Qy 254 KAMLVCS--DA-ATKNFNRLQDVELLPDPSCQWRDTRVYGVFSPWN-Y--SAVCVYSL 307
Db 343 SDIQTVFNGPFAHKEGPNHQLIPYGRIPYPRGTCPGAFTPNMRTTKEFPDDVVTFIR 402
Qy 308 GDIDKVFRTS-SLK-G-----YHSSLPNPRPGKLPDQ-QP-IPET-F--QVADRH 353
Db 403 NHPLMYNPIYTHKRPLIIRIGTDYKTKIAVDVRYNAADGRC-HVLFLGTDQGTQKVVV 461
Qy 354 EVAQRVEPMGPL-KTPLF-H--SKYHOKVAVHRMQASHGETFHVLYLTDRGTHIKVVE 409
Db 462 LPTNFSAGELILEEVLVFSQNSPTTMMKISSKKOOLYVSSSEEGVTQVPLHRCRYGTAC 521
Qy 410 -PGEOEHSAFNIMEIOFRAAAIQTMSLDAERRKLYVSSQWESVQVPLDCEVYGGC 468
Db 522 ADCCCLARDPYCAWDGNSCRYPYTKRRSRQDVRHGNPLTCRCGFNLKAYRNAAEIVQ 581
Qy 469 HGCLMSRDPYCGWDOGRCSISYSE--RSVLQSNPAEPHKECP--NPKPKAPLQKVS 524
Db 582 YGVKNNTTFLECAPKSPQASIKWLQKDN 609
Qy 525 AP-NSRYLSCPMESRHATYSNR-HKEN 550

RESULT 10
ID Q13372; PRELIMINARY: PRT: 753 AA.
AC Q13372;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN III FAMILY HOMOLOG.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 96230324.
RA XIANG R.H., HENSEL C.H., GARCIA D.K., CARLSON H.C., KOK K., DALY M.C.,
RA KERBACHER K., DEN BERG A., VELDHOIS P., BUYS C.H., NAYLOR S.L.;
RT "Isolation of the human semaphorin III/F gene (SEMA3F) at chromosome
3p21, a region deleted in lung cancer.";
RL GENOMICS 32:39-48(1996).
DR EMBL: U38276; G1061351; -.
DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 753 AA; 84941 MW; BECBEBB0 CRC32;
```


Qy 139 PFSPDENSLVLFEGDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNPOFIK 194
Db 261 AELIPDSAEERNDKLYFFFRERSAE-APOSAPAVYARIGICLNDDGGHCCL-VNKWSTFL 318
Qy 195 ATIVHOD-QAYDDKIYFFREDNPDKNFEAPLNVSRAQLCRGOGGESSLSVSKWNTFL 253
Db 319 KARLVCSYVPGEDGIETHDELQDVEV-QOTQDV-RNPVIYAVFTSSGSVFRGSACVYISM 376
Qy 254 KAMLVCS--DAATNKNFRLQDVFLLPDPGQWRDTRVGVESNP-WNY--SACVYISL 307
Db 377 ADIRVFNPGFAHKEGPNYQWMPFSGKMPYPRPGTCPCGGTTPSMKSTKYDPDEVINPMR 436
Qy 308 GDIDKVERTS-SLK-G--YH--SS-LPNRPCKCLPDQ-QP-I-PTETF--QVAD--R 351
Db 437 SHPLMYQAVYPLQRRPLVVRT-GAPYRLTTIAVDQVDAGDG-RYEVFLGTDRGTQVKVI 494
Qy 352 -HPEVAQRVEPMG--PLKTPLEHSHYHVKYAVHMQASHGETFHVLYLTDRGTIHKV 408
Db 495 VLPKDDQEMEELMUEVEVFKDPAPVKMTTSSKRQOQLYVASAVGVTHLSLHRQOAYGAA 554
Qy 409 E-PGEQHSFAFIMEIQPERRAAAIQTMSLDAERRKLYVSSQWESQVPLDLCEVYGG 467
Db 555 CADCCCLARDPYCANDGQACSRYTASSKRRSRQDVYRHGNPIRQCRGENSNANKNAVESVQ 614
Qy 468 CHGCLMSRDPYCGWDQGRCS-YYSSER-SVLQISINPAEPHKECP--NPKPDKAPLOKVS 523
Db 615 YGVAGSAAFLECPQPSQOATVKW 637
Qy 524 LA-PNSRYILSCPWESRHATYSW 545

Search completed: Thu Jul 8 19:35:12 1999
Job time : 48 secs.

W P S R E L

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:40:06 1999; Maspar time 28.37 Seconds
Tabular output not generated. 437.841 Million cell updates/sec

Title: >US-09-041-236-2

Description: (1-584) from US09041236.pap (38 of 45)

Perfect Score: 4352

Sequence: 1 LLLLLWAAAASAGQLRSGP.....ILFIENLTAQQYGHVCEAQ 584

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.465; Variance 152.609; scale 0.239

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	656	15.1	771	13	Human semaphorin III	5.90e-51
2	615	14.1	477	13	Human collapsin.	6.46e-47
3	612	14.1	775	33	Human semaphorin.	1.28e-46
4	606	13.9	861	22	Mouse CD100 antigen.	4.98e-46
5	599	13.8	861	32	Human semaphorin.	2.42e-45
6	544	12.5	776	32	Rat semaphorin W.	5.90e-40
7	538	12.4	862	22	Human CD100 antigen.	2.27e-39
8	463	10.6	441	13	Vaccinia virus semaph	4.37e-32
9	459	10.5	730	13	Grasshopper semaphori	1.06e-31
10	428	9.8	587	32	Human semaphorin W.	1.03e-28
11	390	9.0	974	33	Human secreted protei	4.44e-25
12	378	8.7	712	13	Tribolium semaphorin	6.15e-24
13	367	8.4	650	13	Drosophila semaphorin	6.81e-23
14	362	8.3	724	13	Drosophila semaphorin	2.02e-22
15	352	8.1	930	32	Human semaphorin Y.	1.78e-21
16	325	7.5	888	25	Human semaphorin Z.	6.16e-19

17	316	7.3	887	25	W19856	Rat semaphorin Z.	4.27e-18
18	296	6.8	929	32	W57259	Rat semaphorin Y.	3.09e-16
19	157	3.6	122	13	R71385	Varola major virus S	6.88e-04
20	107	2.5	111	32	W51315	Human semaphorin W pa	6.20e+00
21	99	2.3	591	4	R23006	Protein transcribed f	2.37e+01
22	98	2.3	832	33	W61092	Taq DNA polymerase I	2.80e+01
23	101	2.3	2329	25	W25038	Partial BRCA2 cancer	1.70e+01
24	96	2.2	832	33	W61090	Taq DNA polymerase I	3.88e+01
25	91	2.1	478	1	R04881	Recombinant elastase.	8.70e+01
26	91	2.1	598	9	R48631	Sequence of nuclear r	8.70e+01
27	90	2.1	638	5	R03924	E. coli HSP (dnak).	1.02e+02
28	92	2.1	832	33	W61091	Taq DNA polymerase I	7.41e+01
29	92	2.1	854	30	W56309	Clas II S-receptor ki	7.41e+01
30	92	2.1	855	6	R29815	S receptor kinase pro	7.41e+01
31	93	2.1	4572	30	W52845	A. mediterranei rifam	6.32e+01
32	89	2.0	534	25	W25031	Partial BRCA2 cancer	1.19e+02
33	89	2.0	554	13	R66209	Novel thermostable DN	1.19e+02
34	89	2.0	680	4	R23143	Mutant thermostable D	1.19e+02
35	89	2.0	832	13	R76693	DNA-polymerase (F73L	1.19e+02
36	89	2.0	832	1	P90556	Purified native therm	1.19e+02
37	89	2.0	832	33	W61087	Taq DNA polymerase I	1.19e+02
38	89	2.0	832	33	W61088	Taq DNA polymerase I	1.19e+02
39	89	2.0	832	13	R76690	Taq DNA-polymerase RE	1.19e+02
40	89	2.0	832	33	W61089	Taq DNA polymerase I	1.19e+02
41	89	2.0	833	36	W59942	Amino acid sequence o	1.19e+02
42	89	2.0	833	27	W24211	Cleavage DN nuclease.	1.19e+02
43	89	2.0	833	27	W24212	Cleavage DA nuclease.	1.19e+02
44	89	2.0	3418	26	W19211	Human breast cancer s	1.19e+02
45	89	2.0	3418	26	W23287	Human breast and ovar	1.19e+02

ALIGNMENTS

RESULT 1
ID R71380 standard; Protein; 771 AA.

AC R71380;
DT 21-NOV-1995 (first entry)
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW Varola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 60-63; 101pp; English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87446) or varola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA;

Query Match 15.1%; Score 656; DB 13; Length 771;
Best Local Similarity 31.2%; Pred. No. 5.90e-51;
Matches 143; Conservative 115; Mismatches 155; Indels 45; Gaps 35;

Db 166 rgksppgkltasllldgelysgta-adfmgrodfairtlghhhprtcdhsrwndp 224


```
Db 311 dvfllp-trdp-kopvifglnttsnifrgavcuyhmssireafngpyahkegpeyhws 368
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 273 DVFLLPDSGQWRDRVYGVFSNPWN-Y-S-ACVYSLGDDIKVETRS-SLK-G--YHSS 325
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 369 lyegkvpyprrgscaskngkygttkdpddairfarmhplmyqpkpvhkkipilvktld 428
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 326 L-----PNPRPGKCLPD-QQPI--PTETFF-QVADRHPEVAQRV-EPMGPL-KTP-LFHS- 372
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 gkynlrqlavdrveaedq-ydvflfigtdtgiavlkvitynqetwmeevleelqifkd 487
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 373 -KYHYQKVAHVMQASHGETHVLVLTDRGTIHKVVPGEQHSFAFN-IME-IQPFRR 429
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 papliismeisskrqllyigsasavaqvrffhcdmvggsacacclardycawdgiscsry 547
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 430 AAAQTMSLDAERKLYVSSQWEVSQVPLDICEVYGGCHGCLMSRDPYCGWDQRCISI 489
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 ypt 550
: : :
Qy 490 YSS 492

RESULT 4
ID W17658 standard; Protein: 861 AA.
AC W17658;
DE 24-JUL-1997 (first entry)
DE Mouse CD100 antigen.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
KW vaccine.
OS Mus sp.
FH Key
FT peptide 1..41 Location/Qualifiers
FT /label= Sig_peptide
FT /label= Mat_protein
FT /label= Mat_protein
FT /label= Semaphorin_domain
FT /label= Semaphorin_domain
FT /label= Ig-like_domain
FT /label= Stalk_domain
FT /label= Stalk_domain
FT /label= Transmembrane_domain
FT /label= Transmembrane_domain
FT /label= Cytoplasmic_domain
FT /label= Cytoplasmic_domain
FT modified_site 807..814
FT /label= Phosphorylation
FT /note= "putative tyrosine phosphorylation site"
PN W05717368-A1.
PD 15-MAY-1997.
PF 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND ) DANA FARBER CANCER INST.
PI Boussiottis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB; T60666.
FT Nucleic acid molecule encoding Cd100 antigen - which stimulates
FT leukocyte response, e.g. B cell aggregation, differentiation,
FT survival and T cell proliferation
PS Example 8; Page 86-89; 135pp; English.
CC Mouse CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60666) isolated from murine T cells. Human CD100 antigen
CC (W17657) has also been identified. CD100 polypeptides and fusion
CC proteins, nucleic acids, and host cells expressing CD100 can be
CC utilised in diagnostic and therapeutic methods involving modulation
CC of B and T cell responses, neuron axonal growth and immune cell-
CC nerve cell interaction.
SQ Sequence 861 AA;

Query Match 13.9%; Score 606; DB 22; Length 861;
Best Local Similarity 32.4%; Pred. No. 2.42e-45;
Matches 143; Conservative 105; Mismatches 146; Indels 47; Gaps 31;
```

```
Best Local Similarity 32.4%; Pred. No. 4.96e-46;
Matches 143; Conservative 106; Mismatches 145; Indels 47; Gaps 31;
Db 107 eclnyirvlqlstslvycnfnafqptcdhnltsfklgsedgkrcpfdpahsyts 166
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 93 DCENYITLLER-RSEGLLACGTNARHPSG--WNLVNGTVVPLGEM-RGYAPFSPDENSIV 148
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 167 vmvgelysgts-yNFLgseplisnsshsplrteyaipwlnepsfvadvlqkspdpge 225
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 149 LFEGDEVYSTIRKQYNGKIPFRFRIRGESELYTSDTV--MONPOFIKATIVHQ--D--Q 202
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 geddkvyyffftevsveyefvfkmlprvarvckgqgglrtlq-kkwtstfkarlicskp 284
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 203 AYDDKIYFFREDPNPKNEAPLVNSRVAQLCRDGGOGSESLSVSKWTFKLKMLVCSDA 262
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 dsglvfnllqdvflrap-gl-kepvayvftpqlnnvglsavcaytlatveavfsrky 342
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 263 ATKNFNRLQDVFLLPDPDSGQWRDRVYGVFSNPWNY---SACVYSLGDDIKVF-R--- 315
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 mqsatveqshkwwryngvptprgacidsaaraanytsslnlpdkltlqfkdhplmdd 402
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 316 --TSSL-----K-G-YHSSLNPRPGKCL-PDQQPIP-TETFOVADRHPE-V-AQR-VE- 360
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 svtpidnprkllkdvnyvtqivvdtqalgtfydvfmfistdrgalhhkaviltkev--v 460
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 361 PMGPKLT-P-LFHSKYHYQKVAHVMQASHGETHVLVLTDRGTIHKVVPGEQHSFA 418
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 461 --ieetqlfrdseplvtllsskkgkrfyagsgvvgaplafcekhgs-cedcvlard 517
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 419 FNIMEIQPFRAAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDICEVYGGCHGCLMSRD 476
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 518 pycawspaikacvtlhqeeas 538
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 477 PYCGWDQG-R-CISYSSERS 495

RESULT 5
ID W58540 standard; Protein: 861 AA.
AC W58540;
DE 02-SEP-1998 (first entry)
DE Human semaphorin.
KW Human; semaphorin; diagnosis; nervous disease; immune disease;
KW nerve extension inhibiting activity.
OS Homo sapiens.
PN J10155490-A.
PD 16-JUN-1998.
PF 27-NOV-1996; 332900.
PR 27-NOV-1996; JP-332900.
PA (SUMO ) SUMITOMO SEIYAKU KK.
DR WPI: 98-391044/34.
DR N-PSDB; V31121.
FT New human semaforin gene - useful in the diagnosis of nervous system
FT and immune disorders
PS Claim 1; Page 10-12; 15pp; Japanese.
CC The present sequence represents human semaphorin (translated from the
CC Japanese specification as semaforin). Semaphorin has nerve extension
CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
CC treatment and researches on nervous diseases and immune diseases.
SQ Sequence 861 AA;

Query Match 13.8%; Score 599; DB 32; Length 861;
Best Local Similarity 32.4%; Pred. No. 2.42e-45;
Matches 143; Conservative 105; Mismatches 146; Indels 47; Gaps 31;
Db 107 eclnyirvlqlstslvycnfnafqptcdhnltsfklgsedgkrcpfdpahsyts 166
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 93 DCENYITLLER-RSEGLLACGTNARHPSG--WNLVNGTVVPLGEM-RGYAPFSPDENSIV 148
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 167 vmvgelysgts-yNFLgseplisnsshsplrteyaipwlnepsfvadvlqkspdpge 225
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 149 LFEGDEVYSTIRKQYNGKIPFRFRIRGESELYTSDTV--MONPOFIKATIVHQ--D--Q 202
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 geddkvyyffftevsveyefvfkmlprvarvckgqgglrtlq-kkwtstfkarlicskp 284
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```



```
QY 29 HVGODRVD-FCQTEPHTVLFH--EPGSSSVVVGGRKVYLFDFPEGKN-ASVRTV-NI-G 82
Db 92 ahrelcykkseddcnyirvakiidddrllcgtnaykplchrhvalkdqdyvvekeye 151
QY 83 STKGSC-L-DKR--DCENYITLLERSEG--LLACGTNARHPSCWN--LVNGT-VVPLIG-E 133
Db 152 grglcfdpdmnstaiyseqlysatv-adfsgtdp-li-yrg-p-lrtrsd1-k-qln 204
QY 134 MRGYAPSPDENSLVLEFGEDEVSTIRKQYNGKIPFRFRIRGESELYTSDTVNQPOFI 193
Db 205 apfnv-tmeyndafiffretaveyincgkaly-srvarcvkhdkgg-phqggdrwtsf 261
QY 194 KATIVHODAYDDKIYFFREDNPD-KNPEAPLNVSRVAQLCRGDQGESLSVSKWNTF 252
Db 262 lkarlncsvpgdyfyfneiqstsdiegn99q-vekllygyfttpvnsiggsavcafs 320
QY 253 LKAWLVCSDAATNK-NFNRLQDVF-LLPDP-SGOWDRTRYGVFESPNW-Y--SAVCVYS 306
Db 321 mksilesfdgpfkeqetmnsnwlavpslkyppeprpgcqvndstlpcdvsnfvkshlmd 380
QY 307 L-G-D-IDKVRT-S5LKG-YHS--SL--PNRPCKCLPDQQPIPTETFOVADRHPEVA 356
Db 381 eavpafft-rpilirslqyrfkikavdqvrtpdgdqkaydvlfigtddgkvikalnsaf 439
QY 357 QRVEPMGLKTPLFH-S-KYHQKVAV-HRMAQSHGETFHVLYLTDRGTIHKVVEPG-- 411
Db 440 dsstdvsvvleelqlppgvpvknlyvrmddgddsklvvssdellaiklhrcgsdkit 499
QY 412 EQEHSFAFNIME-IQPFRRAAAIQTM---SLDAERRKLYVSSQWEVSQVPLDLCEVYG-G 466
Db 500 ncrecvslqdpycawdneikctavgspdw 530
QY 467 GCHGCLMSRDPYCGWQD-G-RCISIIYSERS 495

RESULT 10
ID W51314 standard; Protein; 587 AA.
AC W51314;
DE 08-SEP-1998 (first entry)
KW Human; semaphorin W;
KW Human; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW Immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PN WO9815638-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMO) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-PSDB: V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 9.8%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 1.03e-28;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgddelyffftetsrafsdyerikvprvarvcagdlgrkrlq-qrwttfkadllcp 115
QY 201 DQAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGDQGESLSVSKWNTFLKMLVCS 260
Db 116 gpehgrassvlqdvavlrpelga-gtpifygfssqvegatisavcafrpqdirtvlnpg 174
```

```
QY 261 DAATNKNENRLQDVFLLPDPSPGOWDRTRYGVFESPNW---YSAVCVYSLGDDKVFRTS 317
Db 175 frelkhdnrglpvvdndvqprrpgecitanmklrhfgsslsldpdrvtlfrchpmdrp 234
QY 318 --SLK-GYHSSLP--N---PRPGKCLPDQQIP--TETFOVADRHPE-V-AQR-VE-P 361
Db 235 vfpadghp1llvtttdtaylvvahrvtlsqgkydvlylgtedghlbravrigaq-ls-v1 292
QY 362 MGPLKT-PLFHSK-YHQKAVAVHRMOASHGETFHVLYLTDRGTIHKVVEPGQEHSAF 419
Db 293 edlal--fpepgqvemk1-yhsw-llvgsrtevtgtvntncgrlgs-cseciladpvc 347
QY 420 NIMEIQPFRRAAAIQTMSLDAERRKLYVSSQWEVSQVPLDLCEVYGGCHGCLMSRDPYC 479
Db 348 awsfrldecvahagehrglvqdesavslcpegrpvpfvpvataahvvlpccps 407
QY 480 GWD-Q-GRCISIIYSERSVLSQINPAEPHKCPNPKAPLQKVS LAPNSRYLSCPME 537
Db 408 sawascvwhqpsgv 421
QY 538 SRHATYSWRHKENV 551

RESULT 11
ID W64221 standard; Protein; 974 AA.
AC W64221;
DE 06-OCT-1998 (first entry)
KW Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN WO9827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GENY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-362774/31.
DR N-PSDB: V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17j; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 9.0%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 4.44e-25;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechrfikvllknddalfvcgtnafnpscrnykndtlpfdgdfsgmarcpdydahan 172
QY 91 KRDCENVIT-LLERRSEGLLACGTNARHPSCWNLVNGTVVPLG-EMRGYA--PFSPDENS 146
Db 173 valfadgklysatv-dflaidaviyrs1-gesp--tlrtvkhdkwlkpepyfvqadvyg 228
QY 147 LVLFEGDEVYSTIRKQYNG-KIPFRFRIRGESELYTSDTVMONPOFIKATIVHQDQAYD 205
```



```

QY 318 -SLKG-Y-H-S-LPNRPCKCLP-DQPIPTETFEQVADRHE-V-AQRV-EP-MGELK 366
Db 411 hqpllttsralltqvavdgmagphsn-itvmflgsndgtvlkvltpggrsggpepille 469
QY 367 -TPLFH--SKYHYOKVAVHRMQASHGETFHVLYLTIDRGTIHKVVEPGEQEHSAFNIME 423
Db 470 eidaysparcsgkrtaqtarriigleldteghrlfvaafsgciyvlpisrcarhga-cqrs 528
QY 424 -IQPF---R---RAA--A--IQTMSLDAERRKLYVSSQWESQVPLDLCEVYGGGCH-G 470
Db 529 clasqdpvcgwhsrgcvgdirsggtdvdqagngesmehgdcqd 572
QY 471 CLMSRDPYCGWDQGR-CISIYSSE-RSVLQSQINPAE-PHKECPN 511

```

Search completed: Thu Jul 8 19:40:43 1999
Job time : 37 secs.

THIS PAGE BLANK (USPTO)

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:39:15 1999; Maspar time 26.52 seconds
882.213 Million cell updates/sec
Tabular output not generated.

Title: >US-09-041-236-2
Description: (1-584) from US09041236.pap (38 of 45)
Perfect Score: 4352
Sequence: 1 LLLLLMAAASAGHLRSGP.....ILFIENLTAQQYGHYFCEAQ 584

Scoring table: PAM 150

Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60

1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 49.165; Variance 86.170; scale 0.571

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	1875	43.1	653	2 T03102	semaphorin homolog A3	0.00e+00
2	737	16.9	748	2 I48744	semaphorin A - mouse	2.31e-135
3	725	16.7	751	2 I48748	semaphorin E - mouse	1.40e-132
4	716	16.5	749	2 G01856	semaphorin V - human	1.69e-130
5	688	15.8	753	2 G02173	semaphorin III family	5.00e-124
6	681	15.6	772	2 A49069	collapsin - chicken	2.06e-122
7	669	15.4	772	2 I48747	semaphorin D - mouse	1.20e-119
8	662	15.2	666	2 I58169	semaphorin III - mouse	4.90e-118
9	656	15.1	771	2 D49423	semaphorin III precursor	1.18e-116
10	618	14.2	782	2 I48746	semaphorin C - mouse	6.13e-108
11	522	12.0	834	2 S66498	M-sena F protein precursor	3.96e-86
12	474	10.9	760	2 I48745	semaphorin B - mouse	2.31e-75
13	470	10.8	403	2 E42521	A39R protein - vaccin	1.80e-74
14	463	10.6	441	2 S29921	hypothetical protein	6.52e-73
15	461	10.6	730	2 JH0798	fasciclin IV precursor	1.82e-72
16	448	10.3	1074	2 JC5928	semaphorin F precursor	1.40e-69
17	389	8.9	711	2 A49423	semaphorin I precursor	1.30e-56
18	367	8.4	656	2 B49423	semaphorin I - fruit	7.67e-52
19	362	8.3	724	2 C49423	semaphorin II precursor	9.19e-51
20	293	6.7	295	2 J01775	Sall9R protein - vacc	3.91e-36
21	196	4.5	1884	2 JC4975	plexin 2 precursor -	8.84e-17
22	170	3.9	1872	2 JC4976	plexin 3 precursor -	5.47e-12
23	168	3.9	1894	2 JC4980	plexin 1 precursor -	1.25e-11

24	171	3.9	1905	2 J51553	Plexin - African claw	3.62e-12
25	157	3.6	122	2 JQ1845	l4R protein - variola	1.08e-09
26	157	3.6	122	2 H36852	A43R protein - variol	1.09e-09
27	155	3.6	142	2 JQ1776	SaIR1R protein - vacc	2.43e-09
28	120	2.8	775	2 E70320	polyribonucleotide nu	1.28e-03
29	105	2.4	227	2 S09922	hypothetical protein	1.97e-01
30	105	2.4	235	1 Q0BEC9	HXLF4 protein - human	1.97e-01
31	103	2.4	406	2 B69064	conserved hypothetical	3.74e-01
32	102	2.3	275	2 A34866	T-cell surface protei	5.10e-01
33	99	2.3	275	2 S08464	T-cell alloantigen RT	1.29e+00
34	99	2.3	460	2 S35772	translation elongatio	1.29e+00
35	99	2.3	563	2 S78224	virulence-associated	1.29e+00
36	99	2.3	591	2 S26565	virulence-associated	1.29e+00
37	99	2.3	591	2 S03498	virulence-associated	1.29e+00
38	99	2.3	591	2 S22664	virulence-associated	1.29e+00
39	99	2.3	593	2 S15215	virulence-associated	1.29e+00
40	99	2.3	597	2 A46050	thyroid/steroid recep	1.29e+00
41	98	2.3	641	2 S32017	flagellum-associated	1.75e+00
42	98	2.3	970	2 C57282	ankyrin-related prote	1.75e+00
43	98	2.3	979	2 B57282	ankyrin-related prote	1.75e+00
44	99	2.3	1375	2 JC5148	hepatocyte growth fac	1.29e+00
45	98	2.3	1786	2 A57282	ankyrin-related prote	1.75e+00

ALIGNMENTS

RESULT 1
ENTRY #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS T03102
REFERENCE 214840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-653 ##label ENS
##cross-references EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY #length 653 #molecular-weight 73645 #checksum 5501

Query Match 43.1%; Score 1875; DB 2; Length 653;
Best Local Similarity 46.3%; Pred. No. 0.00e+00;
Matches 272; Conservative 105; Mismatches 197; Indels 13; Gaps 13;

Db	57	LLMIL-SAITAAKSRFDKPRILVNLDTGFGQHRF-FGQEPHTVLFHLSNSDYYVGN	114
Qy	1	LLLLLMAAASAGHLRSGPRIFAVMKGVGDVDFGQTEPTVLFHEPGSSVWVGR	60
Db	115	NTIYLFDEAHSSNASTALINITSTHRLSTGCENITLLHNQTDGLLACGTNSQKPS	174
Qy	61	GRVYLFDFPEGNKASVTVNIGSTKGCCLDKRDCENITLLERSEGLLACGTNAHP	120
Db	175	W-LINLNTAQFGLGLGAPFPSPGNSNLVFDQNDTYSTINLYKSLSGS-HKFRRTAG	232
Qy	121	WNLVGTVPV-LGEMRGYAPFPDENSLLVFEQDEYVSTIRK-QEYNGKIPFRIRGE	178
Db	233	ELYTSDTAMHRPQVOATAVHKNSYDDKIYFFFOENSHSDFKQPHTVPRVQVCS	292
Qy	179	ELYTSDTVNQNPQIKATVHQDQYDDKIYFFREDNPKNPAPLNVSRVLAQLCRG	238
Db	293	GESSLSVYKWTFTLKARLACVYDYGRTGRYNELODIFIWQAPENSWEETLI	352
Qy	239	GESSLSVSKWTFTLKMLVLCVSDAATKNFNLQVFLLPDPGQMRDTRVYGVFSP	298
Db	353	FSAVCVFTVKDIDHVFYKTSKLNKHHKLPTRPGQCMKNQHVPTETFOVADRYE	412
Qy	299	YSACVYSLGDDIKVFTSSLAGYHSSLPNPRPGKCLPDQPIPTETFOVADRH	358
Db	413	VYQKNMFPPIQSKYIYTKLLVRYVEYG-GVFWATIFYLLTITKTIHIYVREDS	471

[illegible]

```
QY 304 VYSLGIDIKVFTS-SLK-G--YH--S--S-LPNRPCKCLPDQ-QPI-PTETF--QV-A 349
Db 402 FARSHAMPYNPVPINS-RPIMIKTDVDYQFTQIYVDRVDAEDGQ-YDVMFIGTDIGTVL 459
QY 350 -DR-HPEVAQRVEPMGLKTLPLFHSK--YHYQKVAVHRMQASHGETFHVLYLTDRGTH 405
Db 460 KVSYPKTEWHELEVLLEEMTVFREPTVISAMKISTTQQQLYIGSANGVSQLPHRCDV 519
QY 406 KVE-PGEQHSFAFNIME-IQPFRAAAIQTMSLDAERRKLYSSQWESQVPLDLCEV 463
Db 520 YKACAECCCLARDPYCANDGSSCRYPFTAKRTRRRDIRNGDPLTHCSDLOHHDNPSSQ 579
QY 464 YGGCHGCLMSRDPYCGWDGRCISYSSE-RSVL-QSINPAEPHKECPN-PKPKAP-- 518
Db 580 TLEEKIIYGVENSTFLCSPKSORAIYV-WQFQKONDDHKVE 621
QY 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQSCSE 556

RESULT 7
ENTRY #type complete
TITLE semaphorin D - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48747
REFERENCE I48744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
#cross-references MUID:95267431
#accession I48747
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-772 #label RES
#cross-references EMBL:X85993; NID:g8543329; PID:g854330

GENETICS
#gene semD
#superfamily semaphorin
CLASSIFICATION #length 772 #molecular_weight 88710 #checksum 1776
SUMMARY

Query Match 15.4%; Score 669; DB 2; Length 772;
Best Local Similarity 30.2%; Pred. No. 1.20e-119;
Matches 170; Conservative 143; Mismatches 182; Indels 67; Gaps 51;

Db 58 HTELLDE-ERSRLYVGADHIFSNLVNIKDFQIWPVSY-TRDECKWAGKDLKECA 115
QY 43 HTVLFEHFGSSVWVGKGYLDFDPGKN-AS-VRTVNIIGTKGSC--LDK---RDCE 95
Db 116 NFIKYLEAYNTHLYACGTGAFHPICITYIEYGHHPEDNIFKLQDSHFENGKSPYDPKL 175
QY 96 NYITLERRSEG-LACCTNARHPSCWNL-V-N-GIVVPL--G-EM-RGYAPFSDE 144
Db 176 LTASLLIDGELYSGTA-ADFGNRDFAIFRTLGDHHPITEQHDRLNDPFIASHLIPE 234
QY 145 NSLVFEDEYVSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MONPOFIKATIVHQ 200
Db 235 SDNPDDKRVYFEFRFNAIGGHSKATHARIGQICKNDFGCHRSI-VNKNWTEFKARLIC 293
QY 201 -DQATDDKIYFFREDNPDKNPEAPLVNVRVAQLCRGQGGESSLSVKWTEFLKMLVC 259
Db 294 SVPGPNGIDTHFDELQDVFLM-NSKDP-KNPIVYGVFTTSSNIFKGSVAVC 351
QY 260 S--DA-ATNKNFNRLQDVFLLPDPGQWRDTRVYGVFSNPWN-Y--SAVCYVSLGIDIKV 313
Db 352 FLGPAHRRGNPNQWVPYQGRVPRPCTCPKSTFGGFDSTKLDPDVITGRSHPAWYN 411
QY 314 FRTS-SLK-G--YH--S--S-LPNRPCKCLPDQ-QPI-PTETF--QV-A-DR-HPEVAQ 357
Db 412 PVEFINN-RPIMIKTDVNYQFTQIYVDRVDAEDGQ-YDVMFIGTDVGNLVKVSVPKETW 469
QY 358 RVEPMGLPKTFLFHS--KIHYQKVAVHRMQASHGETFHVLYLTDRGTHIKVVE-PGEQE 414
```

```
Db 470 HDLEEVILLEEMTVFREPTTISAMELSTKQQQLYIGSTAGVAQLPLHRCIDIYKACAECL 529
QY 415 HSAFNIME-IQPFRAAAIQTMSLDAERRKLYSSQWESQVPLDLCEVYGGCGCLM 473
Db 530 ARDPYCAWDGSSCRYPFTAKRTRRRDIRNGDPLTHCSDLEHDHNNHGPSLEERIIYGV 589
QY 474 SRDPYCGWDGRCISYSSE-RSVL-QSINPAEPHKECPN-PKPKD--AP-LQ-KVSLA- 525
Db 590 ENSSTELCSPKSORALVY-WQ 610
QY 526 PNSRYLSC-PMESRHATYSWR 546

RESULT 8
ENTRY #type fragment
TITLE semaphorin III - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I58169
REFERENCE I58169
#authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;
#journal Tessler-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#title Neuron (1995) 14:949-959
#cross-references MUID:95267432
#accession I58169
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-666 #label RES
#cross-references GB:L40484; NID:g703189; PID:g703190

GENETICS
#gene SemIII
#superfamily semaphorin
CLASSIFICATION #length 666 #checksum 9654
SUMMARY

Query Match 15.2%; Score 662; DB 2; Length 666;
Best Local Similarity 30.7%; Pred. No. 4.90e-118;
Matches 141; Conservative 122; Mismatches 149; Indels 48; Gaps 37;

Db 60 RGKSPYDKLLTASLLIDGELYSGTA-ANFMGRDFAIFRTLGHHPHIRTQHDRLNDP 118
QY 135 RGAPFSPDENSILVFEDEGVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNP 190
Db 119 RFIASHLIPSDNPEDDKVYFFFRFNAIDGHSKATHARIGQICKNDFGCHRSI-VNKNW 177
QY 191 QFIKATIVHQ-DQAYDDKIYFFREDNPDKNPEAPLVNVRVAQLCRGQGGESSLSVSKW 249
Db 178 TTFELKARLICSVPGPNGIDTHFDELQDVFLM-NSKDP-KNPIVYGVFTTSSNIFKGSVAVC 235
QY 250 NTFELKMLVCS--DA-ATNKNFNRLQDVFLLPDPGQWRDTRVYGVFSNPWN-Y--SVC 303
Db 236 MYSMSDVRVLLGPAHRRGNPNQWVPYQGRVPRPCTCPKSTFGGFDSTKLDPDVIT 295
QY 304 VYSLGIDIKV---F--RTSS-LK--GYHSSLNPRGPKCLPDQ-QPI-PTETF--QV-A 349
Db 296 FARSHAMPYNPVPINN-RPIMIKTDVNYQFTQIYVDRVDAEDGQ-YDVMFIGTDIGTVL 353
QY 350 -DR-HPEVAQRVEPMGLKTLPLFHS--KIHYQKVAVHRMQASHGETFHVLYLTDRGTH 405
Db 354 KVSVPKETWHDLEEVILLEEMTVFREPTTISAMELSTKQQQLYIGSTAGVAQLPLHRCDI 413
QY 406 KVE-PGEQHSFAFNIME-IQPFRAAAIQTMSLDAERRKLYSSQWESQVPLDLCEV 463
Db 414 YKACAECCCLARDPYCANDGSSCRYPFTAKRTRRRDIRNGDPLTHCSDLOHHDNHGP 473
QY 464 YGGCHGCLMSRDPYCGWDGRCISYSSE-RSVL-QSINPAEPHKECPN-PKPKD--AP 518
Db 474 SLEERIIYGVENSTFLCSPKSORALVY-WQFQRRNEDR 512
QY 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQ 553
```

REFERENCE
#authors
#journal
#title

QY	204	YDKIIYFFREDPNKNEAPLNVSRVAQUCRGDQGGESSLSVSKWNTFLKAMLVCSDA	263
Db	298	-QLPFNIIRHAVLLPADS-P-SYSRIYAVFTSQWVGGRSSAVCAFSJTDIERVEFKGY	354
QY	264	TNKNFNRLQDVFLLPDPGOWRDRTRYGVGSFNPNY-----SACVYSLGDIKVFRT-316	
Db	355	KEINKETSRWTTTVRGSEVSPRPGSCMGSPSSDKALTF-MKD-HFLMDEHVVG-PL---L	408
QY	317	SSL-K-G-----YHSSLNPNRPCKCLPDDQOPIPTETFOVADRHPVAQRVEPMGPLKPL	369
Db	409	VKSGVEYTRLAVESARGLDGSSHWVYLGTSCTGLHKAVVP--QDSS-AFLVPEIOLSPD	465
QY	370	PHSKYHKQVAVRMQAQSHGETHVLVLTDRCTIHKVVEPGQEHSFAFNIHQPFRR	429
Db	466	SEPVNRLQAPAGAVFAGFGSGIWRVRANCSSVYES-CVDCVLARDPHCAMDPESRLCS	524
QY	430	AAAIQTMSLDAERRKLYVSSOWEVSQVPLDLCVYGGCHGLMSRDPYCGWD-QGR-CI	487
Db	525	LLSGSTKPKWKQDMERNPEWCTRGPMARSPRQSPOLIKELVLTVPNSILELRCPHLSA	584
QY	488	SIYSSERSVYLQINPAEPHKECP-NP-K-PDK-AP--LQKVSLA-PNSRYLLSCPMESR	539
Db	585	LASYHWSHGKAKISEA	600
QY	540	HATYSNRH-KENVEQS	554
RESULT	13		
ENTRY		E42521	#type complete
TITLE		A39r protein - vaccinia virus (strain Copenhagen)	
ORGANISM		#formal_name vaccinia virus	
#note		host Homo sapiens (man)	
DATE		09-Nov-1990	#sequence_revision 09-Nov-1990
		08-Apr-1994	#text_change
ACCESSIONS		E42521	
REFERENCE		A33172	
#authors		Johnson, G.P.	
#submission		submitted to GenBank, June 1990	
#accession		E42521	
##status		preliminary	
##molecule_type		DNA	
##residues		length 403	##label JOH
SUMMARY		length 403	##molecular_weight 45741
			##checksum 8167
Query Match		10.8%	Score 470; DB 2; Length 403;
Best Local Similarity		33.5%	Pred. No. 1.80e-74;
Matches		86; Conservative 59;	Mismatches 95; Indels 17; Gaps 16;
Db	77	LVCGTNNGNPKWK-DGSDPKRGRGYAPYQNSKVTTISYN-ECVLSIDINISK-EG-I	132
QY	109	LACGTNARHPSCNWLVNGTVVPLGEMRGYAPFSDENSLVLFEGDEVYSTIRKQYENGKI	168
Db	133	KWRREDPCGYDLYTADNVIKDG-LRGAFTVDKDTYD-KVYILFTDTIGSKR--I-VK	187
QY	169	PRFRIRGE-S-ELYTSDTVNQNPQFIKATIVHQDAYDDKIYFFREDNPDKNPEAPLN	226
Db	188	IPYIAQCLNDEGGPSLSHRHNSITLKVLEEC-DDID-GRSY-R-Q-IHSRTIKTD-ND	241
QY	227	VSRVAQUCRGDQGGESSLSVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLLPDPGQWRD	286
Db	242	TIIVVFPDPSYKSAICTYSMNTIKOSFTSKLEGYTKPLSPAPGICLPAGKVVSHTF	301
QY	287	TRYGVGSFNPNWYSAVCVYSLGDIKVFRTSSLKGYHSSLNPNRPCKCLPDDQOPIETE	346
Db	302	EVIEKYNVLDDIIKPLS	318
QY	347	QVADRHPVAQRVEPMG	363
RESULT	14		
ENTRY		S29921	#type complete
TITLE		hypothetical protein 15 - vaccinia virus	

THIS PAGE BLANK (USPTO)

QY 131 LGE-MRGYAPFSPDENSESLVLFEGDEVYSTRKQYNGK-IPFRRIIRGESE-LYTSQVDM 187
Db 252 ES 253
QY 188 QN 189

RESULT 6
ID NRT1_RAT STANDARD; PRT; 275 AA.
AC P17982;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 (T-CELL MONO(ADP-
DE NAD(P)(+)-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL SURFACE PROTEIN
DE RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN
DE RT6.1).
GN ARTZA OR RT6-A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:LEWIS A;
RX MEDLINE; 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell
RT alloantigen RT6.1";
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
RN [2]
RP MUTAGENESIS OF GLN-207.
RX MEDLINE; 96275529.
RA MAHAMA T., HOSHINO S.-I., KATADA T.;
RT "Increase in ADP-ribosyltransferase activity of rat T lymphocyte
RT alloantigen RT6.1 by a single amino acid mutation.";
RL FEBS LETT. 388:189-191(1996).
CC -!- FUNCTION: HAS NAD+ GLYCOHYDROLASE ACTIVITY AND EXTREMELY LOW ADP-
CC RIBOSYLTRANSFERASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
CC N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -!- CATALYTIC ACTIVITY: NAD(+) + H(2)O = NICOTINAMIDE + ADP-RIBOSE.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -!- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52082; G57168; -.
CC EMBL; M31138; G206804; -.
CC PIR; S08464; S08464.
CC PROSITE; PS01291; ART; 1.
CC PFAM; PF01129; ART; 1.
CC TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
KW T-CELL DIFFERENTIATION; GPI-ANCHOR.
FT SIGNAL 1 20
FT CHAIN 21 246
FT PROPEP 247 275
FT T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
FT HYDROPHOBIC, REMOVED DURING MATURATION
FT (BY SIMILARITY).
FT GPI-ANCHOR (BY SIMILARITY).
FT LIPID 246 246
FT ACT_SITE 209 209
FT CARBOHYD 58 58
FT POTENTIAL.
FT MUTAGEN 207 207
FT Q->E: INCREASED ADP-RIBOSYLTRANSFERASE
FT ACTIVITY.
SQ SEQUENCE 275 AA; 3138 MW; 523B1A84 CRC32;

Query Match 2.3%; Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 3.21e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;
Db 194 LGVYIKFES-FYDQEE-VLIPGYEYQVKTQGYNEIFLDSPKRKSNNCLYSSAGTR 251
QY 131 LGE-MRGYAPFSPDENSESLVLFEGDEVYSTRKQYNGK-IPFRRIIRGESE-LYTSQVDM 187
Db 252 ES 253
QY 188 QN 189

RESULT 7
ID EF1A_TREIR STANDARD; PRT; 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEF1.
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:QM9414 / RUTC-30;
RX MEDLINE; 94124021.
RA NAKARI T., ALATALO E., PENTTILA M.;
RT "Isolation of trichoderma resei genes highly expressed on glucose-
RT containing media: characterization of the tef1 gene encoding
RT translation elongation factor 1 alpha.";
RL GENE 136:313-318(1993).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z23012; G312887; -.
CC PIR; S35772; S35772.
CC PROSITE; PS00301; EFATOR_GTP; 1.
CC PFAM; PF00009; GTP_EFTU; 1.
CC HSP; P0157; IAIP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 49830 MW; 59ABAB2 CRC32;
Query Match 2.3%; Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 3.21e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;
Db 210 YGKEKETKAGKFTG-KTLLEAISIPEPKR-PTDKPLRLPLQDV 252
QY 478 YCGWDQGRCTSIYSERSVLSQINPAEPKPCPNKPKDAPLQKV 522

RESULT 8
ID VRP2_SALTY STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

```
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN MKAA.
OS SALMONELLA TYPHIMURIUM.
OG PLASMID 96 KB VIRULENCE PEX102.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-TML R66;
RX MEDLINE: 90136009.
RA TAIRA S., RHEN M.;
RT "Identification and genetic analysis of mkaA -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth.";
RL MICROB. PATHOG. 7:165-173(1989).
[2]
RP SEQUENCE OF 1-10.
RX MEDLINE: 91244138.
RA TAIRA S., BAUMANN M., RIIKONEN P., SUKUPOLVI S., RHEN M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium.";
RL FEMS MICROBIOL. LETT. 61:319-323(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z15042; G47783; -.
DR PIR: A54540; A54540.
DR PLASMID: VIRULENCE.
KW DOMAIN 367 373 POLY-PRO.
FT SEQUENCE 591 AA; 65440 MW; B69F25C3 CRC32;
SQ
Query Match 2.3%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.21e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLGKTAARLSDPOAASHTAQW-LVEESVTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RDCENYITLLERRSEGLACGTNARHPSCNVLNVTVPVPLGEMRGYAPFSPDENSVLVFE 151
Db 220 GNEAGRDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPFRFRIRGESELY 181
RESULT 10
ID VRP2_SALCH STANDARD; PRT; 591 AA.
AC P17450;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN SPVB.
OS SALMONELLA ENTERITIDIS.
OG PLASMID PNL2001.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-AL1190;
RX MEDLINE: 94362897.
RA SUZUKI S., KOMASE K., MATSUI H., ABE A., KAWAHARA K., TAMURA Y.,
RA KIJIMA M., DANBARA H., NAKAMURA M., SATO S.;
RT "Virulence region of plasmid PNL2001 of Salmonella enteritidis.";
RL MICROBIOLOGY 140:1307-1318(1994).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D14490; G517164; -.
DR PLASMID: VIRULENCE.
KW DOMAIN 367 373 POLY-PRO.
FT SEQUENCE 591 AA; 65350 MW; 9C5B2EE3 CRC32;
SQ
Query Match 2.3%; Score 99; DB 1; Length 591;
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN MKAA.
OS SALMONELLA TYPHIMURIUM.
OG PLASMID 96 KB VIRULENCE PEX102.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-TML R66;
RX MEDLINE: 90136009.
RA TAIRA S., RHEN M.;
RT "Identification and genetic analysis of mkaA -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth.";
RL MICROB. PATHOG. 7:165-173(1989).
[2]
RP SEQUENCE OF 1-10.
RX MEDLINE: 91244138.
RA TAIRA S., BAUMANN M., RIIKONEN P., SUKUPOLVI S., RHEN M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium.";
RL FEMS MICROBIOL. LETT. 61:319-323(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z15042; G47783; -.
DR PIR: A54540; A54540.
DR PLASMID: VIRULENCE.
KW DOMAIN 367 373 POLY-PRO.
FT SEQUENCE 591 AA; 65440 MW; B69F25C3 CRC32;
SQ
Query Match 2.3%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.21e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLGKTAARLSDPOAASHTAQW-LVEESVTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RDCENYITLLERRSEGLACGTNARHPSCNVLNVTVPVPLGEMRGYAPFSPDENSVLVFE 151
Db 220 GNEAGRDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPFRFRIRGESELY 181
RESULT 9
ID VRP2_SALCH STANDARD; PRT; 591 AA.
AC P17450;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN (PROTEIN M2, IN MBA REGION).
OS SALMONELLA CHOLERAE-SUIS (SALMONELLA ENTERICA).
OG PLASMID PKDSC50.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-RF-1;
RX MEDLINE: 90245675.
RA MATSUI H.;
RT "Nucleotide sequences of genes encoding 32 kDa and 70 kDa
RT polypeptides in mba region of the virulence plasmid, PKDSC50, of
```

Best Local Similarity 23.3%; Pred. No. 3.21e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHAQW-LVEESTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RDCENYITLLRRSEGLLACTNARHPSCWNLVNGTVVPLGEMRGYAPFSPDENSILVFE 151
Db 220 GNEAGRDSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPRRIRGESELY 181

RESULT 11

ID VRP2_SALDU STANDARD; PRT: 593 AA.
AC P24419;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN VSDC.
OS SALMONELLA DUBLIN.
OG PLASMID PSDL2.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91251759.
RA KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINEY D.;
RT "Molecular analysis of the virulence locus of the Salmonella dublin
Plasmid PSDL2";
RL MOL. MICROBIOL. 5:307-316(1991).
CC -1- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
OF SALMONELLAS.
CC -1- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
PLASMIDS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: X56727; G47839; -
DR PIR: S15215; S15215.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 375 POLY-PRO.
SQ SEQUENCE 593 AA; 65613 MW; A7CE57B5 CRC32;

Query Match 2.3%; Score 99; DB 1; Length 593;
Best Local Similarity 23.3%; Pred. No. 3.21e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHAQW-LVEESTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RDCENYITLLRRSEGLLACTNARHPSCWNLVNGTVVPLGEMRGYAPFSPDENSILVFE 151
Db 220 GNEAGRDSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPRRIRGESELY 181

RESULT 12

ID PNA1_AJECA STANDARD; PRT: 916 AA.
AC Q07421;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).
GN PNA1
OS AJELLOWYCES CAPSULATA (HISTOPLASMA CAPSULATUM).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;

OC ONYGENALES; ONYGENACEAE; AJELLOWYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94124018.
RA SCHAPER M.P., DEAN G.E.;
RT "Cloning and sequence analysis of an H(+)-ATPase-encoding gene from
the human dimorphic pathogen Histoplasma capsulatum.";
RL GENE 136:295-300(1993).
CC -1- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING
EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKINIZATION MAY MEDIATE
GROWTH RESPONSES.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
(E1-E2 ATPASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: L07305; G409249; -
DR PROSITE: PS00154; ATPASE_E1_E2; 1.
DR PFAM: PF00122; E1-E2_ATPase; 1.
KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
KW ATP-BINDING.
FT DOMAIN 1 111 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 112 134 1 (POTENTIAL).
FT DOMAIN 135 136 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 137 155 2 (POTENTIAL).
FT DOMAIN 156 279 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 280 305 3 (POTENTIAL).
FT DOMAIN 306 320 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 321 350 4 (POTENTIAL).
FT DOMAIN 351 711 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 712 734 5 (POTENTIAL).
FT DOMAIN 735 776 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 777 817 6 (POTENTIAL).
FT DOMAIN 818 843 7 (POTENTIAL).
FT TRANSMEM 844 847 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 848 874 8 (POTENTIAL).
FT DOMAIN 875 916 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 933 962 ASP/GLU-RICH (ACIDIC).
FT MOD_RES 374 374 PHOSPHORYLATION (BY SIMILARITY).
FT BINDING 470 470 ATP (BY SIMILARITY).
SQ SEQUENCE 916 AA; 98884 MW; 4A468A44 CRC32;

Query Match 2.3%; Score 99; DB 1; Length 916;
Best Local Similarity 37.2%; Pred. No. 3.21e-01;
Matches 16; Conservative 8; Mismatches 16; Indels 3; Gaps 3;

Db 762 LAVGTWIT-LTTLMLVSENGGIVQNEGRTHPVLFLEISLTENW 803
QY 244 LSVSKWNTFLKAMLVCS-D-AATNKNFNRLQDV-FLLPDPSGQW 284

RESULT 13

ID DNBI_HSV6U STANDARD; PRT: 1132 AA.
AC F52358;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MAJOR DNA-BINDING PROTEIN (MDBP).
GN U41
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
BETAHERPESVIRINAE; ROSELOVIRUS.


```

RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE: 95266321.
RA GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution.";
RL VIROLOGY 209:29-51(1995).
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES' DNA-BINDING PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X92436; G1044871; -.
CC EMBL: X83413; G854020; -.
CC PFAM: PF00747; Viral_DNA_bp.1.
CC DNA-BINDING; DNA REPLICATION; ZINC-FINGER; NUCLEAR PROTEIN.
CC ZN_FING 459 475
CC SEQUENCE 1132 AA; 127762 MW; DB699A02 CRC32;
CC -----
Query Match 2.3%; Score 98; DB 1; Length 1132;
Best Local Similarity 28.2%; Pred. No. 4.52e-01;
Matches 24; Conservative 18; Mismatches 37; Indels 6; Gaps 6;
Db 435 WNLNRMVYNAGNATEI-YNHLVNCANLCEFCGKC-C-QSCIGTAMVRVGTPLPAIP 491
QY 297 WNYSAVCYVSLGDIKVFRTSLKGYHSSLPNRPCKLDPQOPIPTETQVADRHPVA 356
Db 492 KNVK-KEPLVMSMF-SRY-YAEVDI 513
QY 357 QRVEPMGPKLPLFHSKYHQVAV 381
-----
RESULT 14
ID HA21 HUMAN STANDARD; PRT; 255 AA.
AC PA1908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84168117.
RA AUFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPES D.,
RA STROMINGER J.L.;
RT "Isotypic and allotypic variation of human class II
RT histocompatibility antigen alpha-chain genes.";
RL NATURE 308:327-333(1984).
RN [2]
RP SEQUENCE OF 40-255 FROM N.A.
RX MEDLINE: 83065171.
RA AUFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT "cDNA clone for the heavy chain of the human B cell alloantigen DC1:
RT strong sequence homology to the HLA-DR heavy chain.";
RL PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC -----
RN 112
RP SEQUENCE FROM N.A.
RX MEDLINE: 95266321.
RA GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution.";
RL VIROLOGY 209:29-51(1995).
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES' DNA-BINDING PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC -----
CC EMBL: X92436; G1044871; -.
CC EMBL: X83413; G854020; -.
CC PFAM: PF00747; Viral_DNA_bp.1.
CC DNA-BINDING; DNA REPLICATION; ZINC-FINGER; NUCLEAR PROTEIN.
CC ZN_FING 459 475
CC SEQUENCE 1132 AA; 127762 MW; DB699A02 CRC32;
CC -----
Query Match 2.3%; Score 98; DB 1; Length 1132;
Best Local Similarity 28.2%; Pred. No. 4.52e-01;
Matches 24; Conservative 18; Mismatches 37; Indels 6; Gaps 6;
Db 435 WNLNRMVYNAGNATEI-YNHLVNCANLCEFCGKC-C-QSCIGTAMVRVGTPLPAIP 491
QY 297 WNYSAVCYVSLGDIKVFRTSLKGYHSSLPNRPCKLDPQOPIPTETQVADRHPVA 356
Db 492 KNVK-KEPLVMSMF-SRY-YAEVDI 513
QY 357 QRVEPMGPKLPLFHSKYHQVAV 381
-----
RESULT 15
ID RBSR_ECOLI STANDARD; PRT; 329 AA.
AC P25551;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOSE OPERON REPRESSOR.
GN RBSR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6 AND 250-256.
RC STRAIN-K12;
RX MEDLINE: 93278299.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural and functional analyses of the repressor, RbsR, of the
RT ribose operon of Escherichia coli.";
RL PROTEIN SCI. 1:831-842(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 93315143.
RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL GENOMICS 16:551-561(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE: 93278300.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural homology between rbs repressor and ribose binding protein
RT implies functional similarity.";
RL PROTEIN SCI. 1:843-849(1992).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RBSDACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING

```

```
CC THE RBS OPERON TRANSCRIPTIONAL START SITE. THE AFFINITY FOR THE
CC RBS OPERATOR IS REDUCED BY ADDITION OF RIBOSE, CONSISTENT WITH
CC RIBOSE BEING THE INDUCER OF THE OPERON.
CC -!- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M13169; G147517; -.
CC EMBL: L10328; G290603; ALT_INIT.
CC EMBL: AE000452; G1790194; -.
CC EMBL: D10466; G471110; -.
CC PIR: A41828; A41828.
CC ECOGENE: EG10819; RBSR.
CC PROSITE: PS00356; HTH_LACI_FAMILY; 1.
CC PFAM: PF00356; lacI; 1.
CC PFAM: PF00532; Peripla_BP_like; 1.
CC HSP: P15039; lprv.
CC TRANSCRIPTION REGULATION; REPRESSOR; DNA-BINDING.
CC INIT_MET 0
CC DNA_BIND 3 22 H-T-H MOTIF (BY SIMILARITY).
CC CONFLICT 11 12 GV -> L (IN REF. 1).
CC SEQUENCE 329 AA; 36480 MW; 424E72A6 CRC32;
CC
CC Query Match 2.2%; Score 94; DB 1; Length 329;
CC Best Local Similarity 34.6%; Pred. No. 1.73e+00;
CC Matches 18; Conservative 14; Mismatches 14; Indels 6; Gaps 6;
CC
CC Db 109 TLMQKRVDCILLLC-TETHQPS-REIMQRYPTVP-TVVMDWAPFGDSD-LI 156
CC ||::: ||| | |::: ||::: || | | | | | | | | | | |
CC Qy 99 TLLRRSEGLLA-CGTNARHPSCWNLVNG-TVVPLGEMRGYAPFSPDENSIV 148
```

Search completed: Thu Jul 8 19:37:47 1999
Job time : 23 secs.

 (\mathbf{T}^M)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.

© 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 19:38:07 1999; MasPar time 38.29 Seconds
832.502 Million cell updates/sec
Tabular output not generated.

```

Title: >US-09-041-236-2
Description: (1-584) from US09041236.pep (38 of 45)

```

Perfect score: 4332
Sequence: 1 LLLLWAAAASAOGHLRSGP.....ILFIENLTAOOYGHYFCEAO 584

```

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

sptrembl9

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 48.778; Variance 74.710; scale 0.653

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match	Count					
1	4352	100.0	666	4	075326		SEMAPHORIN L.	0.00e+00
2	2398	55.1	393	11	088371		SEMAPHORIN L (FRAGMENT	0.00e+00
3	1875	43.1	653	14	064906		SIMILAR TO GENBANK ACC	0.00e+00
4	737	16.9	748	11	062177		SEMAPHORIN A PRECURSOR	2.30e-155
5	725	16.7	751	11	062181		SEMAPHORIN E PRECURSOR	3.50e-152
6	716	16.5	749	4	013214		SEMAPHORIN V.	8.50e-150
7	716	16.5	750	4	093018		SEMAPHORIN V.	8.50e-150
8	710	16.3	751	4	099985		SEMAPHORIN E.	3.30e-148
9	705	16.2	751	13	042236		COLLAPSSIN 3.	6.94e-147
10	688	15.8	753	4	013372		SEMAPHORIN III FAMILY	2.17e-142
11	682	15.7	754	11	088633		SEMAPHORIN IV ISOFORM	8.32e-141
12	682	15.7	785	11	088632		SEMAPHORIN IV ISOFORM	8.32e-141
13	681	15.6	772	13	090607		COLLAPSSIN.	1.53e-140
14	681	15.6	785	4	013275		SEMAPHORIN IV.	1.53e-140
15	679	15.6	785	4	015704		SEMAPHORIN.	5.15e-140
16	669	15.4	772	11	062180		SEMAPHORIN D PRECURSOR	2.23e-137
17	664	15.3	772	11	063548		SEMAPHORIN III/COLLAPS	4.64e-136
18	656	15.1	771	4	014563		SEMAPHORIN-III.	5.93e-134
19	618	14.2	775	11	070275		SEMAPHORIN H.	5.63e-124
20	618	14.2	782	11	062179		SEMAPHORIN C (SEM C) (5.63e-124

21	599	13.-8	861 11	009126	SEMAPHORIN J (SEMAPHORIN KIA00331.	5.25e-119
22	596	13.-7	775 4	015041	COLLAPSN-2.	3.19e-118
23	577	13.-3	761 13	Q90663	COLLAPSN-3 (FRAGMENT)	2.88e-113
24	568	13.-1	294 13	Q90664	COLLAPSN-5.	6.32e-111
25	572	13.-1	785 13	Q42337	SEMAPHORIN.	5.76e-112
26	538	12.-4	862 4	Q92854	SEMAPHORIN I (M-SEMA F	3.82e-103
27	522	12.-0	834 11	Q64151	COLLAPSN-5 (FRAGMENT)	5.20e-99
28	511	11.-7	299 13	Q90666	COLLAPSN-4 (FRAGMENT)	3.53e-96
29	473	10.-9	295 13	Q90665	SEMAPHORIN B PRECURSOR	1.97e-86
30	474	10.-9	760 11	Q62178	FASCICLIN IV.	1.09e-86
31	461	10.-6	730 5	Q26473	SEMAPHORIN F HOMOLOG.	2.26e-83
32	448	10.-3	1074 4	Q13591	SEMAPHORIN F PRECURSOR	4.55e-80
33	441	10.-1	1077 11	Q62317	SEMAPHORIN G PRECURSOR	2.71e-78
34	416	9.-6	1093 11	Q60519	SEMAPHORIN F (FRAGMENT)	5.53e-72
35	390	9.-0	434 4	Q60408	SEMAPHORIN VIA.	1.79e-65
36	392	9.-0	888 11	Q35464	SEMAPHORIN-I PRECURSOR	5.69e-65
37	378	8.-7	712 5	Q26972	CESEMA.	1.74e-62
38	367	8.-4	562 5	Q17330	SEMAPHORIN-I.	9.22e-60
39	367	8.-4	771 5	Q24322	SEMAPHORIN-1I.	9.12e-60
40	362	8.-3	706 5	Q24323	SEMAPHORIN IV HOMOLOG	1.58e-58
41	341	7.-8	284 11	Q54948	SEMAPHORIN N (SEMAPHORIN Z.	2.28e-53
42	316	7.-3	886 11	Q54951	SIMILAR TO SEMAPHORIN (FRAG	2.70e-47
43	316	7.-3	887 11	Q70141	KIA00463 PROTEIN (FRAG	2.09e-39
44	283	6.-5	770 5	Q44253		2.34e-21
45	204	4.-7	1963 4	Q70501		

ALIGNMENTS

[illegible]

```
Query Match      100.0%; Score 4352; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Db	33	LLLLLWAAAASQAQHLRSGRIFAVKMGHVQDQDVDFGQTETHTVLFHEPBGSSVNWVGR	92
QY	1	LLLLLWAAAASQAQHLRSGRIFAVKMGHVQDQDVDFGQTETHTVLFHEPBGSSVNWVGR	60
Db	93	GKYYLDFDFEGKKNASVRYTNIGSTKGSCLDKRCENYITLLRRSEGLLACGTNARHPSC	152
QY	61	GKYYLDFDFEGKKNASVRYTNIGSTKGSCLDKRCENYITLLRRSEGLLACGTNARHPSC	120
Db	153	WNLVNGTVVPLGBMRGYAPFSPDENSELVLFEGDEVYSTTKQBYNGKGIAPFRIRGESEL	212
QY	121	WNLVNGTVVPLGBMRGYAPFSPDENSELVLFEGDEVYSTTKQBYNGKGIAPFRIRGESEL	180
Db	213	YTSDTVMQNPQFIKATIVHQDQAYDDKIYFFREDPNKPEAPLNVSVAQLCRGDGG	272
QY	181	YTSDTVMQNPQFIKATIVHQDQAYDDKIYFFREDPNKPEAPLNVSVAQLCRGDGG	240
Db	273	ESSLSVSKNNTFLKAMLVCSDAATNKNFNRLQDVFLLPDPSGOWRDTRVYGVFSNPWNYS	332
QY	241	ESSLSVSKNNTFLKAMLVCSDAATNKNFNRLQDVFLLPDPSGOWRDTRVYGVFSNPWNYS	300

```
Db 333 AVCVYSLGDIKVFRTSSLUKGHSSLPNRPCKLPDQOPIPTETFTQVADRHPEVAQRVE 392
|||||
Qy 301 AVCVYSLGDIKVFRTSSLUKGHSSLPNRPCKLPDQOPIPTETFTQVADRHPEVAQRVE 360
|||||
Db 393 PMGPLKTPLFHSHKYHYOKVAVHRMQASHGETFHVLYLTTDRGTIHKVVPEGEQSHSFAFN 452
|||||
Qy 361 PMGPLKTPLFHSHKYHYOKVAVHRMQASHGETFHVLYLTTDRGTIHKVVPEGEQSHSFAFN 420
|||||
Db 453 IMEIOPFRRAAAIQTMSLDAERKKLYVSSQWESVQVPLDLCBVGCGCHCLMSRDPYCG 512
|||||
Qy 421 IMEIOPFRRAAAIQTMSLDAERKKLYVSSQWESVQVPLDLCBVGCGCHCLMSRDPYCG 480
|||||
Db 513 WDQGRCSIYSSERSVLQSIINPAEPHKECPNPKDKAPLQKVS LAPNSRYILSCPMESRH 572
|||||
Qy 481 WDQGRCSIYSSERSVLQSIINPAEPHKECPNPKDKAPLQKVS LAPNSRYILSCPMESRH 540
|||||
Db 573 ATYSWRHKNENVEQCEPHGQSPNCILFLENLTAQOYGHYFCEAQ 616
|||||
Qy 541 ATYSWRHKNENVEQCEPHGQSPNCILFLENLTAQOYGHYFCEAQ 584
|||||

RESULT 2
ID O88371 PRELIMINARY; PRT: 393 AA.
AC O88371;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN L (FRAGMENT).
GN SEMAL.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCUTIROGNATHI; MURIDAE; MURINAE; MUS.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030699; G3523117; -.
FT NONTER 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;

Query Match 55.1%; Score 2398; DB 11; Length 393;
Best Local Similarity 90.4%; Pred. No. 0.00e+00;
Matches 329; Conservative 16; Mismatches 15; Indels 4; Gaps 2;

Db 33 LLLVFWAAAQAQHSRSGPRISAYWKG---QDHVDFSQPEPHTVLPHEPGSFSVWVYVGR 89
|||||
Qy 1 LLLLLLWAAAQAQHLRSGPRIFAVWKGVHGVQDRVDFGQTEPHTVLPHEPGSSVWVYVGR 60
|||||
Db 90 GKVIYHFNPEGKNASVRTVNTGSTKGSCQDQDCGNITILLERGNGLLVCGTNARHPSC 149
|||||
Qy 61 GKVIYLFDPPEGKNASVRTVNTGSTKGSCDRDCENYITLLERSEGLLACGTNARHPSC 120
|||||
Db 150 WNLVNDVSVMSLGMKGYAPSPDENSLVLFEGDGVYSTIRKQEVNGKIPRFRIRGESE 209
|||||
Qy 121 WNLVNGTVV-PLGEMRGYAPSPDENSLVLFEGDGVYSTIRKQEVNGKIPRFRIRGESE 179
|||||
Db 210 LYTSDTMQNPQFIKATIVHQDQAYDDRIYFFFRFEDNPKNPEAPLNVSRYAQLCRGDQ 269
|||||
Qy 180 LYTSDTMQNPQFIKATIVHQDQAYDDRIYFFFRFEDNPKNPEAPLNVSRYAQLCRGDQ 239
|||||
Db 270 GESSLSVSKWNTFLKAMLVCSDAATNPNRLQDVFLLPDPGSGQWRDTRVYGVFSNPWNY 329
|||||
Qy 240 GESSLSVSKWNTFLKAMLVCSDAATNPNRLQDVFLLPDPGSGQWRDTRVYGVFSNPWNY 299
|||||
Db 330 SAVCVYSLGDIKVFRTSSLUKGHSSLPNRPCKLPDQOPIPTETFTQVADRHPEVAQRV 389
|||||
Qy 300 SAVCVYSLGDIKVFRTSSLUKGHSSLPNRPCKLPDQOPIPTETFTQVADRHPEVAQRV 359
|||||
Db 390 EPMG 393
```

```
|||||
Qy 360 EPMG 363
|||||

RESULT 3
ID Q64906 PRELIMINARY; PRT: 653 AA.
AC Q64906;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ACCELAPHINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS500;
RX MEDLINE: 97201573.
RA ENSSER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CS500;
RX MEDLINE: 97404659.
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=CS500;
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; U18243; G1000717; -.
DR EMBL; AF005370; G2337970; -.
SQ SEQUENCE 653 AA; 73645 MW; CCC853C9 CRC32;

Query Match 43.1%; Score 1875; DB 14; Length 653;
Best Local Similarity 46.3%; Pred. No. 0.00e+00;
Matches 272; Conservative 105; Mismatches 197; Indels 13; Gaps 13;

Db 57 LLMIL-SAITAAKSRFTDKPRLIVNLTDGFGQHRF-FGPOEPHTVLPFSLNSSDVYVGN 114
|||||
Qy 1 LLLLLLWAAAQAQHLRSGPRIFAVWKGVHGVQDRVDFGQTEPHTVLPHEPGSSVWVYVGR 60
|||||
Db 115 NTIYLFDFAISSNASTALINITSTHNTLRSLSSCCENFITLLHNOTDGLLACGTNSQRPSC 174
|||||
Qy 61 GKVIYLFDPPEGKNASVRTVNTGSTKGSCDRDCENYITLLERSEGLLACGTNARHPSC 120
|||||
Db 175 W-LINNETTQFLGPKGLAPFSPSSGNLVLFDDQNDYVSTINLYKSLSGS-HKFRRIAGOV 232
|||||
Qy 121 WNLVNGTVV-PLGEMRGYAPSPDENSLVLFEGDGVYSTIRK-QEYNGKIPRFRIRGES 178
|||||
Db 233 ELTSDTMARHPQFOATAYHKNESYDDKIYFFFEQNSHSDFKQFHTVPRVGVCSDDQ 292
|||||
Qy 179 ELTSDTMQNPQFIKATIVHQDQAYDDRIYFFREDNPKNPEAPLNVSRYAQLCRGDQ 238
|||||
Db 293 GESSLSVYKWTFLKARLACVDYDGTGRIYNEIQDIFIQWAPENSWEETLIYGLFLSPWN 352
|||||
Qy 239 GESSLSVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLLPDPGSGQWRDTRVYGVFSNPWN 298
|||||
Db 353 FSACVFTVDXIDHVFKTSKLVNHHKLPTRPGQCKMKNHQPTEFOVADRYPEVADP 412
|||||
Qy 299 YSACVYSLGDIKVFRTSSLUKGHSSLPNRPCKLPDQOPIPTETFTQVADRHPEVAQR 358
|||||
Db 413 VYQKNNAFPFIQSKYIYTKLVYRVEYG-GVFWATIFYLTITIKTIHYVYEDSNSTT 471
|||||
Qy 359 VEPMPGLKTPLFHSHKYHYOKVAVHRMQASHGETFHVLYLTTDRGTIHKVVPEGEQSHF 417
|||||
Db 472 ALNILENPKQAPTONILLDNTNLKLYNSEWSEVSEVPLDLCVYNGDCFCFMSRDP 531
|||||
Qy 418 AFNIMEIQPFRRAAAIQTMSLDAERKKLYVSSQWESVQVPLDLCVYGGCGHCLMSRDP 477
|||||
```

```

Db 532 LCTWNNTC-S-FK-QRVSVETGGPANRTLSEMGCDHVIAPTIVVHQVSIPLLSNSYLSCP 588
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 478 YCGDQGRCSIIYSRSSVLOSINPAEPH-KE-CPNPKPDKAPLOKVSLAPNSRYLSLSCP 535
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 589 AVSNHADYFWTKDGETEKRCVHKTHKNDICILLIANSTTATNGTHVCN 635
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 536 MESRHATYSWRHKNEVOSCEPGHQSPNCILFIENLTAAQQGYHYPCE 582
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 4
ID Q62177 PRELIMINARY; PRT; 748 AA.
AC Q62177;
DT 01-NOV-1998 (TREMBLEL. 08, CREATED)
DT 01-NOV-1998 (TREMBLEL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLEL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN A PRECURSOR (SEM A).
GN SEMAA OR SEMA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAQOA; CHORDATA; VETERBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIURIGOTATHI; MURIDAE; MURINAE; MUS.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=NMRI.
RX MEDLINE; 95267431.
RA PUESCHEL A.W., ADAMS R.H., BETZ H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
RT and creates domains inhibitory for axonal extension.";
RL NEURON 14:941-948(1995).
CC -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC -!- SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.
CC -!- LOW LEVELS FOUND BETWEEN DAYS 10-12.
CC EXPRESSION PEAKS ON DAY 13 WITH MODERATE LEVELS FROM THEN UNTIL
CC BIRTH.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS ONE C2-LIKE DOMAIN.
DR EMBL; X85990; G854324; -.
DR MGD; MGI:107561; SEMAA.
DR PFAM; PF00047; ig: 1.
KW SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;
KW DEVELOPMENTAL PROTEIN.
FT SIGNAL 1 26
FT CHAIN 27 748
FT DOMAIN 586 649
FT DOMAIN 748 AA; 82894 MW; A7E53A8D CRC32;
SQ SEQUENCE 748 AA; 16.9%; Score 737; DB 11; Length 748;
Query Match 16.9%; Score 737; DB 11; Length 748;
Best Local Similarity 33.7%; Pred. No. 2.30e-155;
Matches 168; Conservative 112; Mismatches 166; Indels 52; Gaps 36

Db 112 ECMFVRLHLHAYNTHLLACRTGAFHPTCALWRWATAGTTHASTGPEKLEDGKGKTPYDP 171
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 93 DCENYITLLRRSEG-LIACGTINARHPSC- ---WWLVNGT-V-Y-P--LGMRGVAFESP 142
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 172 RHRPSPVLVGELYSGV-TADLMGRDFTIFRSLGONPSLRTPEHDSRWLNPKFKVKEWI 230
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 143 DENSILVFEQDEVYTIKQEVNKG-IPRFRIRCESELYT-S-DTV-MQNPOFKAT-I 197
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 231 PESENPPDKYITFFRESAVEAAPAMGRMSYRVSGICRNDLGGQRS-L-VNKWTTTFLKAR 289
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 198 VHQOQAYDDKIYFFREDNPKNPE-APLNVSRVAQLCRGQGGESSLSVSKWNTFLKAM 256
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 290 LVCSPVGEQTHFDQLOQVLLS-SR-DROTPLLYAVFTSSGVFGQSAVCVYSMNQVDR 347
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 257 LVCS--DAATNKNFNRLQDVFLLDPDSQGWDRTRYGVFSPWN-Y--SACVYSLGID 311
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 348 RAFLGLPLHKEGPTHQWYSGRVYPRPGMCPKSTFTGFSSTKDFPDVDTQFGNHHPLM 407
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 312 KVFRIS-SLK-G-----YHSSLPNPRGKCLPDQ-QPIP-TETF--QVAD--R-HPEV 355
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 408 YNPVLPMG-G-R-PLFLQVGAGYTTQTIAADRVAADAGH-YDVLFTIGTDVGTVLKVISVPK 464
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Qy	356	AQRPEMGLKTPLF-H--SKYHYQKVAVHRMQASHGETFVHLVLTITDRGTHKHVE-P-	410
Db	465	GRPNSEGLLEELQVFEDSAITSMOISSKRQOLYVASRAVAQIALHRCITALGRACAE	524
Qy	411	GEQHSFAFNMTETQPFRAAAIQTMSLDAERKLIVSSQWESQVLDLCEVGGCGHG	470
Db	525	OCLARDPYCAWDGSACTRFQTAKRRFRRODIRNGDPSTLCSGDSSVLLEKKVLGVES	584
Qy	471	CLMSRDPYCGMDGRCISIIYSERSVL-QSINPAEPHKPCPNKPDKAPLOKQVSLAPNS	528
Db	595	GSAFLECPRLQAHVQW 602	
Qy	529	-RYLLSCPMEISRATYSW 545	
 RESULT 5			
ID	062181	PRELIMINARY; PRT: 751 AA.	
AC	Q62181;		
DT	01-NOV-1998	(TREMBUREL. 08, CREATED)	
DT	01-NOV-1998	(TREMBUREL. 08, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998	(TREMBUREL. 08, LAST ANNOTATION UPDATE)	
DE	SEMAPHORIN E PRECURSOR (SEM E).		
GN	SEMAE OR SEME.		
OS	MUS MUSCULUS (HOUSE).		
OC	EUKARYOTA; METAZOA;		
OC	SCUTOGENATHI; MURIDAE; MURINAE; MUS.		
CC	[1]		
FP	SEQUENCE FROM N.A.		
RC	STRAIN=NMRI;		
RX	MEDLINE; 95267431.		
RA	PUESCHEL A.W., ADAMS R.H., BETZ H.;		
RT	"murine semaphorin D/collapsin is a member of a diverse gene family		
RT	and creates domains inhibitory for axonal extension.";		
RL	NEURON 14:941-948(1995).		
CC	-!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO		
CC	SUBJECT TERRITORIES INACCESSIBLE FOR GROWING AXONS.		
CC	-!- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE).		
CC	-!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.		
CC	MAXIMUM EXPRESSION BETWEEN DAYS 10-12 WITH MODERATE LEVELS FROM		
CC	DAY 13 UNTIL BIRTH.		
CC	-!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.		
CC	CONTAINS ONE C2-LIKE DOMAIN.		
DR	EMBL; X85994; G854332; .		
DR	MGI; MG1:107557; SEMAE.		
DR	PFAM; PF00047; ig; 1.		
KW	SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;		
KW	DEVELOPMENTAL PROTEIN.		
FT	SIGNAL 1 20	POTENTIAL.	
FT	CHAIN 21 751	SEMAPHORIN E.	
FT	DOMAIN 587 649	IG-LIKE C2-TYPE DOMAIN.	
FT	SEQUENCE 751 AA; 85259 MW; B28D6CFE CRC32;		
Qy	Query Match	16.7%; Score 725; DB 11; Length 751;	
Qy	Best Local Similarity	33.0%; Pred. No. 3.50e-152;	
Qy	Matches	148; Conservative 110; Mismatches 149; Indels 41; Gaps 28	
Db	167	SNPNVNVTSVNINEELPSGMVI-DFMGTDAAIERSLTRKMLQRDLQHNSKWLSEPMEVD	225
Qy	139	FSPDENSLVFEGDEVTSIRKEQYNKG-IIPRRIRIGESELVTS--DTV-MQNQPFIK	194
Db	226	AHVIPDGDPNDKAVFYFFFKLERLTDNNRSTKQIHSMIARIENPTGTGORSV-VNKWTTFEL	284
Qy	195	ATIVHQD-QAYDDKIYYFRREDNPDKNEAPLVSRVAQLCRGOGGESSLSVKWNFTL	253
Db	285	KARLVCSVTDGEGPETHFELEDVFL-ETDNP-RTLIVYGIFTSSVSFGKSAVCVYHL	342
Qy	254	KAMLVCS--DA-ATNKNFNRLQDVFLPDPSGQWRDTRVYGFSNPWN-Y--SACVCVYSL	307
Db	343	SDIQTVFNPGPAHEGNPHQLISVQGRIPYPBPCTCGAFTPNNRTTKDPDDVVFTIR	402
Qy	308	GDIDKFVFTS-SLK-G-----YHSSLUPNPRPKCLPDQ-QP-I-PTETF--QVADRRP	353
Db	403	NHPLMYNISPIHRRPLIVRGTYKYTKIAVDRVNAADG-RYHVLFGTDRGTGVOKVVV	461

[illegible]


```
QY 139 PFSPDENSLVLFEGDEVYSTIRKOEYNGK-IPRFRIRGESELYTS--DTV-MONPOFIK 194
Db 261 AELIPDSAERNDDKLYEFFERSAE-AQOSPAYARIGRICLNDDGHCCL-VNKNSTFL 318
QY 195 ATIVHQD-QAYDDKIYFFREDNPKNPEAPLNVSRVAQJCRGQGGESSLSVSKWNTFL 253
Db 319 KARLVCSVPGEDGIETHFDELQDVV-QOTQDV-RNPVIYAVFTSSGVSFSGSVCVYSM 376
QY 254 KAMLVCS--DAATNKNFNLDQVELLPDPGQWRDTRVYGVFESNP-WNY--SACVCYSL 307
Db 377 ADIRMVFNPGFAHKEGPNYOMPFSGKMPYPRGTCGPGTTFPSMKSTKDYDDEVINEMR 436
QY 308 GDIDKVFRTS-SLK-G--YH---SS-LPNRPCKCLPDO-QP-I-PTETF--QVAD--R 351
Db 437 SHPLMVOAVYPLQRRPLVVRT-GAPYRLTTIAVDQVDAGDG-RYEVLFGLTDRGTQKVI 494
QY 352 -HPEVAQRVEPMG--PLKTPLFHSHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVV 408
Db 495 VLPKDDOEMEELMLEEVEVEFKDPAPVKMTITSSKROOLYVASAVGVTHLSLHRCQAYGAA 554
QY 409 E-PGEOHSAFIMEIQPFRAAAIQTMSLDAERRKLYVSSQWEYSQVPLDLCEVYGG 467
Db 555 CADCCCLARDPYCANDGQACSRYTASSKRRSRQDVRHGNPIROCRGFSNANKNAVESVQ 614
QY 468 CHGCLMSRDPYCGWDOGRGIS-IYSSE-SVLOSINPAEPHKECP--NPKPKAPLQKVS 523
Db 615 YGVAGSAAFLECPSPQATVKW 637
QY 524 LA-PNSRYILSCPMSRHATYSW 545
```

Search completed: Thu Jul 8 19:38:55 1999
Job time : 48 secs.

WQ5REH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:43:54 1999; MasPar time 21.88 Seconds
Tabular output not generated. 556.843 Million cell updates/sec

Title: >US-09-041-236-2
Description: (1-573) from US09041236.pap (39 of 45)
Perfect Score: 4257
Sequence: 1 LLLLLWAAASAGHLRSGP.....SCEPGHSPNCILFIENLTA 573

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.264; Variance 148.831; scale 0.244

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	656	15.4	771	R71380	Human semaphorin III	4.47e-52
2	615	14.4	477	R74175	Human collapsin.	5.90e-48
3	612	14.4	775	W63748	Human semaphorin.	1.18e-47
4	606	14.2	861	W17658	Mouse CD100 antigen.	4.72e-47
5	599	14.1	861	W58540	Human semaphorin W.	2.37e-46
6	544	12.8	776	W51313	Rat semaphorin W.	7.41e-41
7	538	12.6	862	W17657	Human CD100 antigen.	2.93e-40
8	463	10.9	441	R71381	Vaccinia virus semaphorin	7.89e-33
9	459	10.8	730	R71379	Grasshopper semaphorin	1.96e-32
10	428	10.1	587	W51314	Human semaphorin W.	2.17e-29
11	390	9.2	974	W64221	Human secreted protei	1.11e-25
12	378	8.9	712	R71384	Tribolium semaphorin	1.62e-24
13	367	8.6	650	R71382	Drosophila semaphorin	1.88e-23
14	362	8.5	724	R71383	Drosophila semaphorin	5.71e-23
15	352	8.3	930	W57260	Human semaphorin Y.	5.26e-22
16	325	7.6	888	W19857	Human semaphorin Z.	2.04e-19

17	316	7.4	887	25	W19856	Rat semaphorin Z.	1.47e-18
18	296	7.0	929	22	W57259	Rat semaphorin Y.	1.16e-16
19	157	3.7	122	13	R71385	Varola major virus s	4.60e-04
20	107	2.5	111	32	W51315	Human semaphorin W pa	5.00e-00
21	101	2.4	2329	25	W25038	Partial BRCA2 cancer	1.40e-01
22	99	2.3	591	4	R23006	Protein transcribed f	1.97e-01
23	98	2.3	832	33	W61092	Taq DNA polymerase I	2.33e-01
24	96	2.3	832	33	W61090	Taq DNA polymerase I	3.26e-01
25	92	2.2	832	33	W61091	Taq DNA polymerase I	6.31e-01
26	92	2.2	854	30	W56309	Class II S-receptor ki	6.31e-01
27	92	2.2	855	6	R29815	S receptor kinase pro	6.31e-01
28	93	2.2	4572	30	W52845	A. mediterranei rifam	5.35e-01
29	91	2.1	478	1	R04881	Recombinant elastase.	7.42e-01
30	89	2.1	534	25	W25031	Partial BRCA2 cancer	1.03e-02
31	89	2.1	534	13	R66209	Novel thermostable DN	1.03e-02
32	91	2.1	598	9	R48631	Sequence of nuclear I	7.42e-01
33	90	2.1	638	5	R03924	E. coli HSP (dnaK).	8.73e-01
34	89	2.1	680	4	R23143	Mutant thermostable D	1.03e-02
35	89	2.1	832	13	R76693	DNA-polymerase (F73L,	1.03e-02
36	89	2.1	832	1	P90556	Purified native therm	1.03e-02
37	89	2.1	832	33	W61087	Taq DNA polymerase I	1.03e-02
38	89	2.1	832	33	W61088	Taq DNA polymerase I	1.03e-02
39	89	2.1	832	13	R76690	Taq DNA-polymerase RE	1.03e-02
40	89	2.1	832	33	W61089	Taq DNA polymerase I	1.03e-02
41	89	2.1	833	36	W59942	Amino acid sequence o	1.03e-02
42	89	2.1	833	27	W24211	Cleavage DN nuclease.	1.03e-02
43	89	2.1	833	27	W24212	Cleavage DA nuclease.	1.03e-02
44	89	2.1	3418	26	W19211	Human breast cancer s	1.03e-02
45	89	2.1	3418	26	W23287	Human breast and ovar	1.03e-02

ALIGNMENTS

RESULT 1
ID R71380 standard; Protein: 771 AA.
AC R71380:
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW varola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN WO9507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 60-63; 101pp; English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II,
CC (Q87444-5), Tribolium semaphorin I (Q87446) or varola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor binding
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA;

Query Match 15.4%; Score 656; DB 13; Length 771;
Best Local Similarity 31.2%; Pred. No. 4.47e-52;
Matches 143; Conservative 115; Mismatches 155; Indels 45; Gaps 35;

Db 166 rgkspdyplktasllidgelysgta-adfmgdrfaifrtlghhphirteqhsrwldnp 224


```

: |||: |||: | : : | : |||: |||: |||: | | : |||: |||:
QY 203 AYDDKIYFFREDNPDKNPEAPLNVSRVAOLCRGDOGESSLSVKWNTFLKAMLVCSDA 262
Db 285 dsglvfnlqdvflrap-gl-kepvfayvftqqlnnvglsavcaytlatveavfgrky 342
QY 263 ATKNNFNRLQDVLFPDPSQWRDTRVYGVSFPWNV---SAYCVYSLGIDIKVF-R--- 315
Db 343 mgsatveqstkwrvyngpytpdpqecidsearaanytsslnlpdkltqgfvkdhplmdd 402
QY 316 --TSSL-----K-G-YHSSLPNRPKCL-POOPIP-TETFOVADRHP-V-AQR-VE- 360
Db 403 svtpidnrpklikdvnvtqivvdrtqalqdtfydmfistdrgalnhkaviltkevnh--v 460
QY 361 PMGPLKT-P-LFHSKYHYQKVAHVHRMQASHGETFHVLYLTDRGTTHKVVPEGEQHSFA 418
Db 461 --leotqlfrdfepvltllsskgrtkrfvyagsngvvgaplafacekhs-cedcvlard 517
QY 419 FNIMEIQPFRAAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDLCEVYGGCHGLMSRD 476
Db 518 pycawspaikacvtllhqeas 538
QY 477 PYCGWOG-R-CISIIYSSERS 495

RESULT 6
ID W51313 standard; Protein; 776 AA.
AC W51313;
DT 08-SEP-1998 (first entry)
DE Rat semaphorin W.
KW Rat; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW Immunosuppressant; gene therapy; diagnosis; research reagent.
OS Rattus norvegicus.
PN W09815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMU ) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-PSDB; V07279, V07280.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Claim 1; Page 60-64; 90pp; Japanese.
CC The present sequence represents rat semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 776 AA;

Query Match 12.8%; Score 544; DB 32; Length 776;
Best Local Similarity 27.4%; Pred. No. 7.41e-41;
Matches 163; Conservative 145; Mismatches 232; Indels 55; Gaps 39;

Db 26 llllllalspvcgrvprvrtslpiseadsyltrfaashtynysallvdpashlyvg 85
QY 1 LLLLLLWAAAASAOGLH-RSGPRI-FAWKGHVGQDRVDFQGTPTHTVLFPEPGSSVWVG 58
Db 86 ardsifaltlpfsgerprridwvmp-ethrcnckrkkgkdedchnfiqilaivnashlht 144
QY 59 GRKQVY-L-EDFPEGKNASV-RVTNVTGSTKGSL--DKR-D-CENITLDERSEG-LLA 110
Db 145 cgtfatdpkcgvdvssfgqverlesgrkcpfepagrsaavmagglytatvk-nflgt 203
QY 111 CGTNARHPSCWNL-VNG--TVVPLGEMRGVAPSPDENSLVLFGEDVYSTIRKQYNGK 167
Db 204 eplisravgraedwritetlsswlnapafvaamvlspeavgdgdgddeiffftetsrvl 263
QY 168 IPRFRIRGESELYT-SDTV---MQNPQFIKATIV-H-Q--DOAYDDKIYFFREDNPDK 219
```

```

Db 264 dsyerikvrvarvcagdlqgrktlq-grwtfflkadllcpgpghgrasvqamaelrp 322
QY 220 NPEAPLNVSRVAOLCRGDOGESSLSVKWNTFLKAMLVCSDAATNKNFNRLQDVLFPD 279
Db 323 qpqa-gtplfyglfssqweaalsavcfrpdqdravlngpfrelkhdcnrglpvmdnev 391
QY 280 PSQWRDTRVYGVSFPWN---YSAVCVYSLGIDIKVFRTS--SLK-GYHSSLP--N-- 328
Db 382 pprpgeciannmkllqfgsslsldrvltfirdhplmdrpfpadgrpllvtttdtaylr 441
QY 329 --PRPGKCLPDQOPIP--TETFOVADRHP-V-AQR-VE-PMGPLKT-PLFHSK-YHYQK 378
Db 442 vvaahrvtlsqkeydvllygtedghlhravrigaq-ls-vledlal--fpeqpvcsmkl 497
QY 379 VAVHRMQASHGETFHVLYLTDRGTTHKVVPEGEQHSFAFNIMEIQPFRAAAIQTMSL 438
Db 498 yndw--llvgshvtqvntscnrglqs-csecilaqdpvcawsfrcidacvahaghrm 554
QY 439 DAERRKLYVSSQWEVSQVPLDLCEVYGGCHGLMSRDPYCGWD-Q-GRCISIIYSSERSV 496
Db 555 vqdiesadvssicpkepgpvhvfvvratvghvlpespsawascvvhqpsgv 609
QY 497 LOSINPAEPHKECPNPKDKAPLQKVLAPNSRYLSCPMSRKHATYSWRHKNV 551

RESULT 7
ID W17657 standard; Protein; 862 AA.
AC W17657;
DT 24-JUL-1997 (first entry)
DE Human CD100 antigen.
KW Human CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
KW vaccine.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..41 /label= Sig_peptide
FT protein 42..862 /label= Ig-like_domain
FT domain 42..553 /label= Mat_protein
FT domain 42..553 /label= Semaphorin_domain
FT domain 554..630 /label= Ig-like_domain
FT domain 631..733 /label= Stalk_domain
FT domain 735..752 /label= Transmembrane_domain
FT domain 753..862 /label= Cytoplasmic_domain
FT modified_site 808..815 /label= Phosphorylation
FT /note= "putative tyrosine phosphorylation site"
PN W09717368-A1.
PD 15-MAY-1997.
PF 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND ) DANA FARBER CANCER INST.
PI Bousiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB; T60665.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation.
PS Claim 7; Page 70-72; 135pp; English.
CC Human CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60665) isolated by COS cell expression cloning. CD100
CC polypeptides and fusion proteins, nucleic acids, and host cells
CC expressing CD100 can be utilised in diagnostic and therapeutic
CC methods involving modulation of B and T cell responses, neuron
CC axonal growth and immune cell-nerve cell interaction.
SQ Sequence 862 AA;
```



```
Oy 29 HVGDRVD--FGQTEPHTVLFH--EPGSSVWVGGRKGYLFDPEBGN-ASVRTV-NI-G 82
Db 92 ahrelcykgseddcqnyrlavkladdrvliqctnqayklcrlhvalkdgdyyvekeye 151
Oy 83 STKGSC-L-DKR--DCENIYLLERSSEG-LLACGTNARHPSCWN--LVNGT-VVPLG-E 133
Db 152 grglcpdpdhnstaiyseglysatv-adfsgtdp-li-yrg-p-lrtersdl-k-qln 204
Oy 134 MRGAPFSPDENSELVLFEDEYVETIRKQYNGKIPRFRIRGESELYTSDTMQNPQFI 193
Db 205 apnfyn-Emeyndfiffretaveyincgaiky-srvarckhdkgg-phgggdrtwsf 261
Oy 194 KATIVHDOAYDDKIIYFFREDNPD-KNPEAPLNVSRAQLCRGDGGESSLVSKWNTE 252
Db 262 lksrlncsvpgdpyfyfeigstsdilegnvggq-vekllygvtftvnsjggsavcafs 320
Oy 253 LKAMLVCSDAATNK-NFNRLODVF-LLPDP-SGQWRDTRVYGVFSPNPN-Y--SAVCVYS 306
Db 321 mksllesfdgpfkceqetmnsnwlavpslkvppeprpgqcvndsrtdpdsvnfvkshtlmd 380
Oy 307 L-G--D-IDKVFRT--SSLKG-YHS--SL--PNRPGKCLPDQOPIPTETFEQVADRHPEVA 356
Db 381 eaapfaft-rpallrisiqy-ftkiavdqqrtpdgkaydvlfigtddgkvikalnsasf 439
Oy 357 QRVEPMGPKLTPLFH-S-KYHYQKAV-HRMOASHGETFHVLYLTDRGTTHKVVPEG-- 411
Db 440 dssdtdvsvleelqvlppgvpnklyvrmddgdsklvvsvsddelaiklhrcgsdkit 499
Oy 412 FOEHSFAFNIME-IOPFRRAAIQTM---SLDAERKLYVSSQWEVSOVPLDLCEVYG-G 466
Db 500 ncrecvsldpvcawndvvelkctavgsdpws 530
Oy 467 GCHGCLMSRDPYCGWDQ-G-RCISIIYSERS 495

RESULT 10
ID W51314 standard; Protein: 587 AA.
AC W51314;
DT 08-SEP-1998 (first entry)
DE Human semaphorin W.
KW Human; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW Immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PN W09815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR (SUMO ) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI; 98-261015/23.
DR N-PSDB: V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;
```

```
Query Match 10.1%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 2,17e-25;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgddeiyffttetsradsverikprvarvcagdlqgrkktlq-qrwttflkadtllcp 115
Oy 201 DOAYDDKIYFFREDNPDKNPEAPLNVSRAQLCRGDGGESSLVSKWNTEFLKMLVCS 260
Db 116 gpehgrassiqdvavrlpelga-gtpifygfssqwegatisavcafrpqqdirtvlnpg 174

Query Match 9.2%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 1,11e-25;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechfnfivllknddalfvcgtnafncpncrykmdtlepfdfgdefsgmarcpydakhan 172
Oy 91 KRDCENYIT-LLERRSGLLACGTNARHPSCWNLVNGTVVPLG-EMRGYA--PFSPOENS 146
Db 173 valfadokklysatvt-dflaidaviyrsf-gesp--clrtvkhdkskwlekpyfqavdyg 228
Oy 147 LVLFEGDEVYSTIRKQBYNG-KIPRFRIRGESELYTSDTMQNPQFIKATIVHQDAYD 205
```

```
RESULT 11
ID W64221 standard; Protein: 974 AA.
AC W64221;
DT 06-OCT-1998 (first entry)
DE Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN W09827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GEM ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 98-362774/31.
DR N-PSDB: V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland CDNA libraries
PS Claim 17j; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain CDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;
```



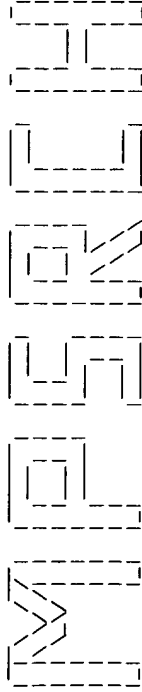
```

QY 318 -SLKG-Y-H-S-S-LPNRPGKCLP-DQQPIPTETFOVADRHP-E-V-AQRV-EP-MGPLK 366
Db 411 hqplltlttsralltqvavdgmagphsn-itvmflgsndgtvkvltpggrsggpepille 469
QY 367 -TPLFH--SKYHYQKVAVHRMQASHGETFHVLYLTDRGTIHKYVEPEQEHSHFAFNIME 423
Db 470 eidaysparcsgkrtagtarriigleldteghrlfvafsgcivylplsrcarhga-cgrs 528
QY 424 -IQPF---R---RAA--A--IQTMSLDAERKLYVSSQWEVSQVPLDLCEVYGGGCH-G 470
Db 529 clasqdpqycgwhsrgcvgdrgsggtdvdqagndesmehdcdq 572
QY 471 CLMSRDPYCGWDQGR-CISIYSSE-RSVLQOSINPAE-PHKECPN 511

```

Search completed: Thu Jul 8 19:44:39 1999
Job time : 45 secs.

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:42:48 1999; MasPar time 24.78 Seconds
Tabular output not generated. 926.544 Million cell updates/sec

Title: >US-09-041-236-2
Description: (1-573) from US09041236.pap (39 of 45)
Perfect Score: 4257
Sequence: 1 LLLLLWAAAASAGHLRSGP.....SCEPGHSPNCILFIENLTA 573

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir50
1.pir1 2.pir2 3.pir3 4.pir4

Statistics: Mean 48.966; Variance 86.740; scale 0.565

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1865	43.8	653	2	T03102		semaphorin homolog A3	0.00e+00
2	737	17.3	748	2	I48744		semaphorin A - mouse	4.87e-134
3	725	17.0	751	2	I48748		semaphorin E - mouse	2.75e-131
4	716	16.8	749	2	G01856		semaphorin V - human	3.18e-129
5	688	16.2	753	2	G02173		semaphorin III family	8.06e-123
6	681	16.0	772	2	A49069		collapsin - chicken	3.20e-121
7	669	15.7	772	2	I48747		semaphorin D - mouse	1.74e-118
8	662	15.6	666	2	I58169		semaphorin III - mouse	6.86e-117
9	656	15.4	771	2	D49423		semaphorin III precursor	1.59e-115
10	618	14.5	782	2	I48746		semaphorin C - mouse	6.76e-107
11	522	12.3	834	2	S66498		M-sema F protein prec	2.63e-85
12	474	11.1	760	2	I48745		semaphorin B - mouse	1.19e-74
13	470	11.0	403	2	E42521		A39R protein - vaccin	9.09e-72
14	463	10.9	441	2	S29921		hypothetical protein	3.18e-72
15	461	10.8	730	2	JH0798		fasciclin IV precursor	8.75e-72
16	448	10.5	1074	2	JC5928		semaphorin F precursor	6.31e-69
17	389	9.1	711	2	A49423		semaphorin I precursor	4.35e-56
18	367	8.6	656	2	B49423		semaphorin I - fruit	2.30e-51
19	362	8.5	724	2	C49423		semaphorin II precurs	2.69e-50
20	293	6.9	295	2	JQ1775		Sall9R protein - vacc	8.22e-36
21	196	4.6	1884	2	JC4976		plexin 2 precursor -	1.22e-16
22	170	4.0	1872	2	JC4976		plexin 3 precursor -	6.84e-12
23	171	4.0	1905	2	I31553		Plexin - African claw	4.54e-12

ALIGNMENTS

RESULT 1
ENTRY #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS T03102
REFERENCE Z14840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-653 #label ENS
#cross-references EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY #length 653 #molecular-weight 73645 #checksum 5501

Query Match 43.8%; Score 1865; DB 2; Length 653;
Best Local Similarity 46.7%; Pred. No. 0.00e+00;
Matches 270; Conservative 104; Mismatches 191; Indels 13; Gaps 13;

Db	57	LLMIL-SAITAAKSRFDIKPLIVNLTGFGQHRF-FGQPEPHTVLFHSLNSSDYVGGN	114
Qy	1	LLLLLWAAAASAGHLRSGPRIFAVKGVGDVDFGOTEPHTVLFHEPGSSVWVGR	60
Db	115	NTIYDFEFAHSSNASTALINITSTHNTLRLSTCENFTILLHNOTDGLLAGTNSQKPS	174
Qy	61	GKVIYDFEPGKNASVTRVNIWGKSGCLDRDCENYITLLERSEGLLAGCTNARHPS	120
Db	175	W-LINLTTQPLKGLAPSPSSGNLVLPDQNTSTINLYKSLGS-HKFRIRAGV	232
Qy	121	WNLVNGITVP-LGEMRGVAPSPDENSILVPEGDEVSTIRK-QEYNGKIRFRIRGES	178
Db	233	ELYTSDTAMHRPQVQATAVHKNSYDDKIYFFQENSHSDFKQPHTPVRVGVCSSDQ	292
Qy	179	ELYTSDTVMQNPQIKATIVHQDAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGQ	238
Db	293	GGESSLYVKTITLTKARLACVDYDTGRIYNELODIFIWQAPENSWEETLYGLFSPWN	352
Qy	239	GGESSLSVKNTEFLKAMLVCSDAATNKNRDLVLLPDPSCQWRDTRVYGFSPWN	298
Db	353	FSAVCVFTVKDIDHVFKTSKLNKYNHKLPTPRPGCKMKNHQHVPTETFOVADRYPEVADP	412
Qy	299	YSACVYSLGIDKVKFRTSSLKGVHSLPNRPCKCLPDQOPIPTETFOVADRYPEVADP	358
Db	413	VIQKNNAFPIQSKIYITKLLVYRVEYG-GVFWATIFYLTITKGTIHYVYEDSNSTT	471

[illegible]

```
QY 304 VYSLGDIKVERTS-SLK-G--YH--S--S-LPNPRGKCLPDQ-QPI-PTETF--QV-A 349
Db 402 FARSHPAMYNPVPINS-RPIMIKTDVDYQFTQIVVDVDRVDAEDGQ-YDVMFICTDGTGL 459
QY 350 -DR-HPEVAORVEPMGLKPTPLFHSK--YHYQKVAHRMQASHGETFHVLYLTDRGTIH 405
Db 460 KVSIPKRETWHEELVEEMTFRPTTISAMKISTKQOOLYIGSAGTQVQLPLHRCDV 519
QY 406 KVEE-PGEQEHSAFNIME-IQFPRRAAIIQTMSLDAERKLYVSSQWESQVPLDLCEV 463
Db 520 YGKACAECLARDPYCAWDGSSCSRYPPTAKRTRRQDINGDPLTHCSDLQHDHNPSSGQ 579
QY 464 YGGGCHGCLMSRDPYCGWDGRCISYSE-RSVL-QSINPAEPHKECPN-PPKDRAP-- 518
Db 580 TLEEKIIYGVNSTECPKSORAIVY-WQFQKNDHDKVE 621
QY 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEOSCE 556

RESULT 7
ENTRY semD #type complete
TITLE semaphorin D mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48747
REFERENCE Puschel, A.W.; Adams, R.H.; Betz, H.
#authors Neuron (1995) 14:941-948
#journal Murine semaphorin D/collapsin is a member of a diverse gene
#title family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48747
#status preliminary; translated from GB/EMBL/DBD
#molecule_type mRNA
#residues 1-772 #label RES
#cross-references EMBL:X85993; NID:9854329; PID:9854330
GENETICS
#gene semD
#superfamily semaphorin
CLASSIFICATION #length 772 #molecular_weight 88710 #checksum 1776
SUMMARY

Query Match 15.7%; Score 669; DB 2; Length 772;
Best Local Similarity 30.2%; Pred. No. 1.74e-118;
Matches 170; Conservative 143; Mismatches 182; Indels 67; Gaps 51;

Db 58 HTFLDLDE-ERSRLVYGAKHIFSNLVNIKDFQKIYVPSY-TRRDECKWAGDKILKECA 115
QY 43 HTVLFEHPGSSSVWVGGRGVYLFDFPEGKN-AS-VRTVNIQSTKGC--LDK---RDCE 95
Db 116 NFKVLKAYNOTHYACGTCAFPICPTYIEVGHHPEDNIFKLODSHFENGSRGKSPYDPKL 175
QY 96 NYITLLERSEG-LLACGTWARPSCWNL-V--N--GTVVPL--G--EM-RGVAFSPDE 144
Db 176 LTASLLDGLYSGTA-ADPMGRDFAIFRTLGDHHPHPIRTEQHDNRWLNDRPFI 234
QY 145 NSLVFEGDEVYTIKQYNGK-IPFRIRGESELYTS--DTV-MQNPFIKATIVHQ 200
Db 235 SDNPEDDKVYFFPRENAIGEGSKATHAIGQICKNDKDFGHRSL-VNKKWTTTLKARLIC 293
QY 201 -DQAYDDKIYFFREDNPDKNPEAPLVSRVAQLCRGDGSGESSLSVSKNTLKA 259
Db 294 SVPGPNIDTHFELODVLFLM-NSKDP-KNPIYGVFTTSSNIFKGSANCYMSDVRV 351
QY 260 S--DA-ATNKNRNLQDVFLLPDPGQWRDTRVYGFSPNWN-Y--SAYCVYSLGDI 313
Db 352 FLGPAHRDGNPQWYQGRVPYPRGTCPSKTFGFGDSTKDLDPDVIITFGRSHPAMYN 411
QY 314 FRIS-SLK-G-YH-S--S-LPNPRGKCLPDQ-QPI-PTETF--QV-A-DR-HPEVAQ 357
Db 412 PVFPINN-RPIMIKTDVNYQFTQIVVDVDRVDAEDGQ-YDVMFICTDGTGLKVVSPK 469
QY 358 RVEPMGLKPTPLFHS--KYHYQKVAHRMQASHGETFHVLYLTDRGTIHKKVVE-PGEQ 414
```

```
Db 470 HDLEEVLLEMTVFRPTTISAMELSTKQOOLYIGSAGTQVQLPLHRCDIYKACAECL 529
QY 415 HGFANIME-IQFPRRAAIIQTMSLDAERKLYVSSQWESQVPLDLCEVYGGGCHGCLM 473
Db 530 ARDPYCAWDGSSCSRYPPTAKRTRRQDINGDPLTHCSDLQHDHNPSSLEERIYGV 589
QY 474 SRDPYCGWDGRCISYSE-RSVL-QSINPAEPHKECPN-PPKDK--AP-LQ-KVSLA- 525
Db 590 ENSSTFLECPKSORALVY-WQ 610
QY 526 PNSRYLSC-PMESRHATYSWR 546

RESULT 8
ENTRY I58169 #type fragment
TITLE semaphorin III - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I58169
REFERENCE Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;
#authors Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#journal Neuron (1995) 14:949-959
#title Semaphorin III can function as a selective chemorepellent to
#cross-references MUID:95267432
#accession I58169
#status preliminary; translated from GB/EMBL/DBD
#molecule_type mRNA
#residues 1-666 #label RES
#cross-references GB:L40484; NID:9703189; PID:9703190
GENETICS
#gene SemIII
#superfamily semaphorin
CLASSIFICATION #length 666 #checksum 9654
SUMMARY

Query Match 15.6%; Score 662; DB 2; Length 666;
Best Local Similarity 30.7%; Pred. No. 6.86e-117;
Matches 141; Conservative 122; Mismatches 149; Indels 48; Gaps 37;

Db 60 RGKSPYDPKLLTASLLDGLYSGTA-ANFMGRDFAIFRTLGHHPHPIRTEQHDNRWLN 118
QY 135 RGYAFSPDENSLVLEEGDEVYSTIRKQYNGK-IPFRIRGESELYTS--DTV-MQNP 190
Db 119 RTISAHLPESNPEDDKVYFFPRENAIDGESHGKATHAIGQICKNDKDFGHRSL-VNKK 177
QY 191 QETKATIVHQ-DQAYDDKIYFFREDNPDKNPEAPLVSRVAQLCRGDGSGESSLSVSKW 249
Db 178 TTFKLKARLICSVPGPNIDTHFELODVLFLM-NSKDP-KNPIYGVFTTSSNIFKGSAYC 235
QY 250 NYTLKAMLYCS--DA-ATNKNRNLQDVFLLPDPGQWRDTRVYGFSPNWN-Y--SAYC 303
Db 236 MYSMSDVRRLVLLGPAHRDGNPQWYQGRVPYPRGTCPSKTFGFGDSTKDLDPDVI 295
QY 304 VYSLGDIKDV----F--RTSS-LK--GYHSSLPNPRPGKCLPDQ-QPI-PTETF--QV-A 349
Db 296 FARSHPAMYNPVPINN-RPIMIKTDVNYQFTQIVVDVDRVDAEDGQ-YDVMFICTDGTGL 353
QY 350 -DR-HPEVAORVEPMGLKPTPLFHS--KYHYQKVAHRMQASHGETFHVLYLTDRGTIH 405
Db 354 KVSVPKRETWHEELVEEMTFRPTTISAMELSTKQOOLYIGSAGTQVQLPLHRCDI 413
QY 406 KVEE-PGEQEHSAFNIME-IQFPRRAAIIQTMSLDAERKLYVSSQWESQVPLDLCEV 463
Db 414 YGKACAECLARDPYCAWDGSSCSRYPPTAKRTRRQDINGDPLTHCSDLQHDHNP 473
QY 464 YGGGCHGCLMSRDPYCGWDGRCISYSE-RSVL-QSINPAEPHKECPN-PPKDK--AP 518
Db 474 SLEERIYGVNSTECPKSORALVY-WQFQRNEDR 512
QY 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQ 553
```


ORGANISM #formal_name vaccinia virus
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997

ACCESSIONS S29921
REFERENCE S29907
#authors Amegadzie, B.Y.
#submission submitted to the EMBL Data Library, January 1991
#accession S29921
#status Preliminary
#molecule_type DNA
#residues 1-441 #label AME
#cross-references EMBL:X57318; NID:g62239; PID:g62254
SUMMARY #length 441 #molecular-weight 50185 #checksum 6034

Query Match 10.9%; Score 463; DB 2; Length 441;
Best Local Similarity 33.9%; Pred. No. 3,18e-72;
Matches 87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;

Db 115 LVCGTNGNPKWK-IDGDDPKHGRGVAPYQNSKVIIHNGC-VLSDINISK-EG-I 170
Qy 109 LACGTNARHPSCNLVNGTVPLGEMRGYAPFSPDENSLVLFEGDEVYSTIRKQYNGKI 168
Db 171 KRRRFDCGCGYLDYADNIPKDG-LRGAFYDKDGYD-KVYILFTDITIGSKR--I-VK 225
Qy 169 PRFRIRGE-S-ELVTSVTQNPQFIRKATIVHQDQAYDDKIYFFREDNPKNPEAPLN 226
Db 226 IPYIAQMLNDEGSSLSHSHWSTFLKVELEC-DID-GRSY-R-Q-IIHSTIKTD-ND 279
Qy 227 VSRVAQLCRGQGGESSLSVSKWNTFLKAMLVCSDAATNKNENLQDVLFLLDPDPSGQWRD 286
Db 280 TILYVFFSPYSKSLCTYKNTIKQSTKLEGYTKOLPSPASSGICLPAGKVYPHTTF 339
Qy 287 TRVGVFNPNWYSAVCVISLGDIDKVFRTSLKGYHSLNPRPKCLPDQOPIPTETF 346
Db 340 EVIEKYNLDDIIPKLS 356
Qy 347 QVADHRPEVAQRVPMG 363

RESULT 15
ENTRY JH0798 #type complete
TITLE fasciclin IV precursor - American bird grasshopper
ORGANISM #formal_name Schistocerca americana #common_name American bird grasshopper
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Sep-1997
ACCESSIONS JH0798
REFERENCE JH0798
#authors Kolodkin, A.; Matthes, D.J.; O'Connor, T.P.; Patel, N.H.; Admon, A.; Bentley, D.; Goodman, C.S.
#journal Neuron (1992) 9:831-845
#title Fasciclin IV: Sequence, expression, and function during growth cone guidance in the grasshopper embryo.
#accession JH0798
#molecule_type mRNA
#residues 1-730 #label KOL
#cross-references GB:I00709; NID:g160844; PID:g160845
#experimental_source embryo
COMMENT This protein plays a role in growth cone guidance in the developing central nervous system.
glycoprotein; transmembrane protein
KEYWORDS
FEATURE
1-22 #domain signal sequence #status predicted #label SIG\
23-730 #product fasciclin IV #status predicted #label MAT\
23-627 #domain extracellular #status predicted #label EXT\
628-652 #domain transmembrane #status predicted #label TMN\
653-730 #domain intracellular #status predicted #label INT\
44,71,163,267,360, #binding_site carbohydrate (Asn) (covalent) #status predicted
539
SUMMARY #length 730 #molecular-weight 81214 #checksum 5881
Query Match 10.8%; Score 461; DB 2; Length 730;

Best Local Similarity 28.0%; Pred. No. 8,75e-72;
Matches 143; Conservative 130; Mismatches 182; Indels 56; Gaps 46;

Db 32 QFGEERVQRLGNESHKDHFKLLEKDHNSLLVGARNIVYNISLRDLTFTQRIEWHSSG 91
Qy 29 HVGQDRVD-FGQTEPHTLFFH--EPGSSSVWVGGRGVYLFDFPESGKN-ASVRTV-NI-G 82
Db 92 AHRELCLYLGKSEDDCQNYIRVLAKIDDDRVLCGTNAYKPLCRHYALKDGVYVVEKEYE 151
Qy 83 STKGSC-L-DKR--DCENYITLLRSEG-LIACGTNARHPSCWN--LVNCT-VVPLG-E 133
Db 152 GRGLCPDPDHNSTAIYSEGOLYSATV-ADFSGTDLP-LI-YRG-P-LRTERSDL-K-OLN 204
Qy 134 MRGVAPFDPDENSLVLFEGDEVISTIRKQYNGKIPRRRIRGSELSLTSDFVMONPOFI 193
Db 205 APNFYN-TMEYNDFIFFFFRETAVIYINCGKAIY-SRVARVCKHDKGPHGFG-DRWTSF 261
Qy 194 KATIVHQDQAYDDKIYFFREDNPD-KNPEAPLNVSRAQLCRGQGGESSLSVSKWNTF 252
Db 262 LKSRLNCVPGDYPFYFNEIOSTSDIIIEGNYGGQ-VEKLIYGVFTTPVNSIGGSVACAFS 320
Qy 253 LKAMLVCSDAATNK-NFNRLQDVF--LLPDP-SGQWRDTRVYGVFSNPWN-Y--SAVCVYS 306
Db 321 MKSILLESFDGPKKEQETMNSNWLAVPSLKVPEPRPGOCVNDSTLDPVSVNFVKSHLTMD 380
Qy 307 L-G--D-IDKVFT--SSLKG-YHS--SL--PNRPGKCLPDQOPIPTETTFQVADHRPEVA 356
Db 381 EAVPAFFT-RPTILIRISLQYRFTKIAVDQVQRTPDGKAYDVLFIGTDDGKVIKALNSASF 439
Qy 357 QRVEPMGLKTPLFH-S-KYHYQKAV-HRMOASHGETFHVLYLTDRGTIHKVVEPG-- 411
Db 440 DSSDTVDSVTEELQVLPPGVPKNLYVVRMDGDDSKLVVVSDDDEILAIIKLHRCGSKIT 499
Qy 412 EOEHSFAFNIME-IQPFRAAAAIQTM---SLDAERKLYVSSQWEVSQVPLDLCEVYG-G 466
Db 500 NCRECVSLQDPYCAWDNVELKCTAVGSPDWS 530
Qy 467 GCHGCLMSRDPYCGWDQ-G-RCISIIYSERS 495

Search completed: Thu Jul 8 19:43:34 1999
Job time : 46 secs.

THIS PAGE BLANK (USPTO)

MAISEH (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 19:41:02 1999; MasPar time 17.99 Seconds
Tabular output not generated. 900.173 Million cell updates/sec

Title: >US-09-041-236-2
Description: (1-573) from US09041236.pep (39 of 45)
Perfect Score: 4257
Sequence: 1 LLLLLWAAASAOQHLSRGP.....SCEPGHSPNCILFIENLTA 573

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 50.181; Variance 76.023; scale 0.660

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	%	Description	Pred. No.
1	470	11.0	403	1 VA39_VACCC PROTEIN A39.	4.36e-86
2	463	10.9	441	1 VA39_VACCV PROTEIN A39.	2.74e-84
3	163	3.8	1871	1 SEX_HUMAN TRANSMEMBRANE PROTEIN H	9.50e-13
4	105	2.5	227	1 US08_HCMVA HYPOTHETICAL PROTEIN H	3.79e-02
5	102	2.4	275	1 NRT2_RAT T-CELL ECTO-ADP-RIBOSY	1.11e-01
6	99	2.3	275	1 NRT1_RAT T-CELL ECTO-ADP-RIBOSY	3.15e-01
7	99	2.3	460	1 EFIA_TIRRE ELONGATION FACTOR 1-AL	3.15e-01
8	99	2.3	591	1 VRP2_SALTY 65 KD VIRULENCE PROTEI	3.15e-01
9	99	2.3	591	1 VRP2_SALCH 65 KD VIRULENCE PROTEI	3.15e-01
10	99	2.3	591	1 VRP2_SALDN 65 KD VIRULENCE PROTEI	3.15e-01
11	99	2.3	593	1 VRP2_SALDU 65 KD VIRULENCE PROTEI	3.15e-01
12	99	2.3	916	1 PMAL_AJECA PLASMA MEMBRANE ATPASE	3.15e-01
13	98	2.3	1132	1 DNBI_HSV60 MAJOR DNA-BINDING PROT	4.44e-01
14	95	2.2	255	1 HAZ1_HUMAN HLA CLASS II HISTOCOMP	1.22e+00
15	94	2.2	329	1 RBSR_ECOLI RIBOSE OPERON REPRESSO	1.70e+00
16	93	2.2	380	1 DP3B_MYCPN DNA POLYMERASE III, BE	2.35e+00
17	92	2.2	442	1 CEMI_YEAST 3-OXOACYL-(ACYL-CARRIE	3.25e+00
18	92	2.2	460	1 EFIA_NEUCR ELONGATION FACTOR 1-AL	3.25e+00
19	95	2.2	518	1 VL2_HPV5B MINOR CAPSID PROTEIN L	1.22e+00
20	95	2.2	518	1 VL2_HPV05 MINOR CAPSID PROTEIN L	1.22e+00
21	95	2.2	683	1 AMOH_ARTGO HISTAMINE OXIDASE (EC	1.22e+00
22	93	2.2	827	1 YZ23_METJA HYPOTHETICAL PROTEIN M	2.35e+00
23	94	2.2	947	1 PMA2_YEAST PLASMA MEMBRANE ATPASE	1.70e+00

24	94	2.2	959	1 N100_YEAST NUCLEOPORIN NUP100/NSP	1.70e+00
25	93	2.2	1029	1 ENDI_YEAST VACUOLAR BIOGENESIS PR	2.35e+00
26	94	2.2	1069	1 ENTK_MOUSE ENTEROPEPTIDASE (EC 3.	1.70e+00
27	94	2.2	1115	1 IREI_YEAST SERINE/THREONINE-PROTE	1.70e+00
28	92	2.2	1132	1 DNBI_HSV62 MAJOR DNA-BINDING PROT	3.25e+00
29	94	2.2	1200	1 DDX8_CAEEL PUTATIVE PRE-MRNA SPLI	1.70e+00
30	95	2.2	1390	1 MET_HUMAN HEPATOCYTE GROWTH FACT	1.22e+00
31	92	2.2	1663	1 CO3_MOUSE COMPLEMENT C3 PRECURSO	3.25e+00
32	90	2.1	119	1 RL19_MYCPN 50S RIBOSOMAL PROTEIN	6.14e+00
33	90	2.1	213	1 AMEX_BOVIN AMELOGENIN, CLASS I PR	6.14e+00
34	90	2.1	254	1 HAZ2_HUMAN HLA CLASS II HISTOCOMP	6.14e+00
35	91	2.1	348	1 DDL_ENTFA D-ALANINE--D-ALANINE L	4.48e+00
36	91	2.1	356	1 VP39_NPVLD MAJOR CAPSID PROTEIN.	4.48e+00
37	90	2.1	380	1 GBB_MAIZE GUANINE NUCLEOTIDE-BIN	6.14e+00
38	89	2.1	404	1 NIFS_ECOLI NIFS PROTEIN HOMOLOG.	8.40e+00
39	89	2.1	437	1 RFBB_MYXXA O-ANTIGEN EXPORT SYSTE	8.40e+00
40	91	2.1	498	1 ELAS_PSEAE PSEUDOLYSIN PRECURSOR	4.48e+00
41	91	2.1	598	1 NOT_HUMAN IMMEDIATE-EARLY RESPON	4.48e+00
42	89	2.1	598	1 RNRI_RAT REGENERATING LIVER NUC	8.40e+00
43	91	2.1	707	1 GCYK_HCMVA GANCICLOVIR KINASE (EC	4.48e+00
44	91	2.1	1666	1 CO3_CAVPO COMPLEMENT C3 PRECURSO	4.48e+00
45	90	2.1	1874	1 POLR_KYMWJ RNA REPLICASE POLYPROT	6.14e+00

ALIGNMENTS

RESULT 1
ID VA39_VACCC STANDARD; PRT; 403 AA.
AC P21062;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE PROTEIN A39.
GN A39R.
OS VACCINIA VIRUS (STRAIN COPENHAGEN).
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91021027.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RT "The complete DNA sequence of vaccinia virus.";
RL VIROLOGY 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RL VIROLOGY 179:517-563(1990).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M35027; G335517; -
DR PIR: E42521; E42521.
SQ SEQUENCE 403 AA; 45741 MW; FE0DB9AB CRC32;

Query Match 11.0%; Score 470; DB 1; Length 403;
Best Local Similarity 33.5%; Pred. No. 4.36e-86;
Matches 86; Conservative 59; Mismatches 95; Indels 17; Gaps 16;

Db 77 LVCGTNGNPKCWK-IDGSDDPKHRGVRGYAQNYSKVITISYN-ECVLSIDNISK-EG-I 132
QY 109 LAGCTNARHPSCNVLNVGVPLGEMGRGYAPSPDENVLFEGDEVYSTIRKQEVNGKI 168
Db 133 KWRREFDPCGVDLYTADNVIPKDG-LRGAFVDKDGTVD-KVYILFTDTIGSKR--I-VK 187
QY 169 PRFRIRGE-S-ELYTSDTVMQNPFIKATIVHQDQYDKKIYFFREDNPKNPEAPLN 226

QY 131 LGE-MRGYAPSPDENSLVLFEGDEVSTIRKQYNGK-IPRFRIRGESE-LYTSVTVM 187
Db 252 ES 253
QY 188 QN 189

Query Match 2.3% Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 3.15e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;

Db 194 LGVYIKFEF-FYPDOEE-VLIPGYEVQKVRTQGYNEIFLDSPKRKSNNCLYSSAGTR 251
QY 131 LGE-MRGYAPSPDENSLVLFEGDEVSTIRKQYNGK-IPRFRIRGESE-LYTSVTVM 187

Db 252 ES 253
QY 188 QN 189

RESULT 6
ID NRT1_RAT STANDARD; PRT; 275 AA.
AC P17982;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 PRECURSOR (EC 2.4.2.31) (T-CELL
DE NAD(P)(+)-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL MONO(ADP-
DE RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN
DE RT6.1).
GN ART2A OR RT6-A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEWIS A;
RX MEDLINE; 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell
RT alloantigen RT6.1".
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
RN [2]
RP MUTAGENESIS OF GLN-207.
RX MEDLINE; 96275529.
RA MAEHAMA T., HOSHINO S.-I., KATADA T.;
RT "Increase in ADP-ribosyltransferase activity of rat T lymphocyte
RT alloantigen RT6.1 by a single amino acid mutation".;
RL FEBS LETT. 388:189-191(1996).
CC -!- FUNCTION: HAS NAD+ GLYCOPOLYMERASE ACTIVITY AND EXTREMELY LOW ADP-
CC RIBOSYLTRANSFERASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
CC N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -!- CATALYTIC ACTIVITY: NAD(+) + H2O = NICOTINAMIDE + ADP-RIBOSE.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -!- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52082; G57168; -.
CC EMBL; M31138; G206804; -.
CC PIR; S08464; S08464.
CC PROSITE; PS01291; ART; 1.
CC PFAM; PF01129; ART; 1.
CC T-CELL DIFFERENTIATION; GPI-ANCHOR.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
FT SIGNAL 1 20
FT CHAIN 21 245 T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
FT PROPEP 247 275 HYDROPHOBIC, REMOVED DURING MATURATION
FT
FT LIPID 246 246 GPI-ANCHOR (BY SIMILARITY).
FT ACT_SITE 209 209 BY SIMILARITY.
FT CARBOHYD 58 58 POTENTIAL.
FT MUTAGEN 207 207 Q>E: INCREASED ADP-RIBOSYLTRANSFERASE
FT ACTIVITY.
SQ SEQUENCE 275 AA; 31388 MW; 523B1A84 CRC32;

Query Match 2.3% Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 3.15e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;

Db 194 LGVYIKFEF-FYPDOEE-VLIPGYEVQKVRTQGYNEIFLDSPKRKSNNCLYSSAGTR 251
QY 131 LGE-MRGYAPSPDENSLVLFEGDEVSTIRKQYNGK-IPRFRIRGESE-LYTSVTVM 187

Db 252 ES 253
QY 188 QN 189

RESULT 7
ID EF1A_TRIE STANDARD; PRT; 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEFL
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-QM9414 / RUTC-30;
RX MEDLINE; 94124021.
RA NAKARI T., ALATALO E., PENTTILA M.;
RT "Isolation of Trichoderma reesei genes highly expressed on glucose-
RT containing media: Characterization of the tef1 gene encoding
RT translation elongation factor 1 alpha".;
RL GENE 136:313-318(1993).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z23012; G312887; -.
CC PIR; S35772; S35772.
CC PROSITE; PS00301; EFACITOR_GTP; 1.
CC PFAM; PF00009; GTP-EFTU; 1.
CC HSSP; P07157; LAIP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 49830 MW; B9ABAB22 CRC32;

Query Match 2.3% Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 3.15e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;

Db 210 YKWEKETAGFTG-KTLLAIDSIETPPKR-PTDKPLRLPLQDV 252
QY 478 YCGWDQGRCSISYSSERSVLQSNPAEPHKECPNPKDPKAPLQKV 522

RESULT 8
ID VRP2_SALTY STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)


```
Best Local Similarity 23.3%; Pred. No. 3.15e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RDCENVITLLRSEGLACGNNARHPSCNVLNGVTVPLGEMRGYAPSPDENSILVLF 151
Db 220 GNEAGDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPRRIRGESELY 181

RESULT 11
ID VRP21 SALDU STANDARD; PRT; 593 AA.
AC P24419;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN VSDC.
OS SALMONELLA DUBLIN.
OG PLASMID PSDL2.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91251759.
RA KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINEY D.;
RT "Molecular analysis of the virulence locus of the Salmonella dublin
plasmid PSDL2."
RL MOL. MICROBIOL. 5:307-316(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X56727; G47839; -
DR PIR: S15215; S15215.
KW DOMAIN 367 375 POLY-PRO.
SQ SEQUENCE 593 AA; 65613 MW; A7CE57B5 CRC32;

Query Match 2.3%; Score 99; DB 1; Length 593;
Best Local Similarity 23.3%; Pred. No. 3.15e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RDCENVITLLRSEGLACGNNARHPSCNVLNGVTVPLGEMRGYAPSPDENSILVLF 151
Db 220 GNEAGDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPRRIRGESELY 181

RESULT 12
ID PMAL_AJECA STANDARD; PRT; 916 AA.
AC Q07421;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).
GN PMAL.
OS AJELLOMYCES CAPSULATA (HISTOPLASMA CAPSULATUM).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;

Best Local Similarity 23.3%; Pred. No. 3.15e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RDCENVITLLRSEGLACGNNARHPSCNVLNGVTVPLGEMRGYAPSPDENSILVLF 151
Db 220 GNEAGDRSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPRRIRGESELY 181

RESULT 13
ID DNBI_HSV6U STANDARD; PRT; 1132 AA.
AC P52338;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MAJOR DNA-BINDING PROTEIN (MDBP).
GN U41.
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; ROSELOVIRUS.
```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95256321.
RA GOMPELUS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution.";
RL VIROLOGY 209:29-51(1995).
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES' DNA-BINDING PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X92436; G104871; -.
DR EMBL; X83413; G854020; -.
DR PFAM; PF00747; viral_DNA_Bp; 1.
KW DNA-BINDING; DNA REPLICATION; ZINC-FINGER; NUCLEAR PROTEIN.
FT ZN_FING 459 475 C4-TYPE.
SQ SEQUENCE 1132 AA; 127762 MW; DB699A02 CRC32;

Query Match 2.3%; Score 98; DB 1; Length 1132;
Best Local Similarity 28.2%; Pred. No. 4.44e-01;
Matches 24; Conservative 18; Mismatches 37; Indels 6; Gaps 6;

Db 435 WNLNRMVYNAGNAHTEI-YNHLVNCNSANLCEFDGKC-C-OSGIGTAMRVGTRLPAP 491
Qy 297 WNSAVCVYSLGDIKIDKFTSLLGYHSLNPRGKCLPDQPIPTFTFQVADRHPVA 356
Db 492 KNVK-KEPLVMSMF-SRY-YAEVDI 513
Qy 357 QRVEPMGPKLPLFHSKYHYQKAV 381

RESULT 14
ID HA21_HUMAN STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84168117.
RA AUFFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPES D.,
RA STROMINGER J.L.;
RT "Isotypic and allotypic variation of human class II
RT histocompatibility antigen alpha-chain genes.";
RL NATURE 308:327-333(1984).
RN [2]
RP SEQUENCE OF 40-255 FROM N.A.
RX MEDLINE; 83065171.
RA AUFFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT "cDNA clone for the heavy chain of the human B cell alloantigen DC1:
RT strong sequence homology to the HLA-DR heavy chain.";
RL PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93278300.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural homology between rbs repressor and ribose binding protein
RT implies functional similarity.";
RL PROTEIN SCI. 1:843-849(1992).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RBSACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J00199; -; NOT_ANNOTATED_CDS.
DR PIR; A02212; HLHUCD.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00993; MHC_II_alpha; 1.
DR HSP; P01910; IIAK.
KW MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 23
FT CHAIN 24 255 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT DQ(1) ALPHA CHAIN.
FT DOMAIN 24 110 EXTRACELLULAR ALPHA-1.
FT DOMAIN 111 204 EXTRACELLULAR ALPHA-2.
FT DOMAIN 205 217 CONNECTING PEPTIDE.
FT TRANSMEM 218 240
FT DOMAIN 241 255 CYTOPLASMIC TAIL.
FT DISULFID 133 189 BY SIMILARITY.
FT CARBOHYD 104 104 POTENTIAL.
FT CARBOHYD 144 144 POTENTIAL.
SQ SEQUENCE 255 AA; 28105 MW; 5C46DD59 CRC32;

Query Match 2.2%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1.22e+00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;

Db 25 DIVADSVAAQLG-VNLYQSYGSPGQSYSHFEDGDEEYVDLERKETWOLPFRFR 78
Qy 122 NLVNGTVPLGEMGYAPFSDENSLVLFEGDEV-YSTIRKOEYNGKIPRRIR 175

RESULT 15
ID RBSR_ECOLI STANDARD; PRT; 329 AA.
AC P25551;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOSE OPERON REPRESSOR.
GN RBSR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6 AND 250-256.
RX STRAIN-K12;
RX MEDLINE; 93278299.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural and functional analyses of the repressor, RbsR, of the
RT ribose operon of Escherichia coli.";
RL PROTEIN SCI. 1:831-842(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE; 93315143.
RA BURKLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL GENOMICS 16:551-561(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RA IIDA A., TESHIBA S., MIZOBUCHI K.;
RT SUBMITTED (XXX-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SIMILARITY TO RIBOSE-BINDING PROTEINS.
RX MEDLINE; 93278300.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural homology between rbs repressor and ribose binding protein
RT implies functional similarity.";
RL PROTEIN SCI. 1:843-849(1992).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RBSACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING


```
Db 333 AVCYSLGDIKVFRTSSLGKHYHSSLPNPRGKCLPDQOPIPTETFOVADRHPEVAQRVE 392
|||||
QY 301 AVCYSLGDIKVFRTSSLGKHYHSSLPNPRGKCLPDQOPIPTETFOVADRHPEVAQRVE 360
Db 393 PMGPLKTPLFHSKYHYQKVAVHRMQASHGETFHVLYLTDDRGTHHKVVEPGEQEHSAFN 452
|||||
QY 361 PMGPLKTPLFHSKYHYQKVAVHRMQASHGETFHVLYLTDDRGTHHKVVEPGEQEHSAFN 420
Db 453 IMEIOPFRRAAATQTHSLDAERKRLYVSSQWVEYSQVPLDLCVEYGGCHGCLMSRDPYCG 512
|||||
QY 421 IMEIOPFRRAAATQTHSLDAERKRLYVSSQWVEYSQVPLDLCVEYGGCHGCLMSRDPYCG 480
Db 513 WDQGRCSISYSSRSVLOSINPAEPHKECPNPKAPLOKVS LAPNSRYLLSCPMESRH 572
|||||
QY 481 WDQGRCSISYSSRSVLOSINPAEPHKECPNPKAPLOKVS LAPNSRYLLSCPMESRH 540
Db 573 ATYSWRHKNVEOSCEPHGOSPCILFIENLTA 605
|||||
QY 541 ATYSWRHKNVEOSCEPHGOSPCILFIENLTA 573

RESULT 2
ID O88371 PRELIMINARY; PRT; 393 AA.
AC O88371;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN L (FRAGMENT).
GN SEMAL.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses.";
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030699; G3523117; -.
FT NON_TER 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;

Query Match 56.3%; Score 2398; DB 11; Length 393;
Best Local Similarity 90.4%; Pred. No. 0.00e+00;
Matches 329; Conservative 16; Mismatches 15; Indels 4; Gaps 2;

Db 33 LLLVFWAAASQAQHSRSGPRISAVWKG---QDRVDFSQEPHTVLFHEPGSFSSVWVGR 89
|||||
QY 1 LLLLLWAAASQAQHLRSRGPRIFAVWKGHVQDRVDFGQTEPTVLFHEPGSSSVWVGR 60
Db 90 GKVFHFNPEGKNASRTVNTVIGSTKGCQKQDCGNTITLLRRGNGLLVCGTNARKPSC 149
|||||
QY 61 GKVYLFDFPECKNASRTVNTVIGSTKGCQKQDCGNTITLLRRSEGLLACGTNARKPSC 120
Db 150 WNLVNDVSVMSLCEMKGYAFSPDENSLVLFEGDEVYSTIRKQYNGKIPFRFRIGESE 209
|||||
QY 121 WNLVNGTVPV-PLGEMRGYAFSPDENSLVLFEGDEVYSTIRKQYNGKIPFRFRIGESE 179
Db 210 LYTSDTVMQNPQFIKATIVHQDAYDDKIYFFREDNPKNPAPLNSVVAQLCRGDQ 269
|||||
QY 180 LYTSDTVMQNPQFIKATIVHQDAYDDKIYFFREDNPKNPAPLNSVVAQLCRGDQ 239
Db 270 GESSLSVSKWNTFLKAMLVCSDAATNFRNLQDVFLLPDPGQWRDTRVYGVFSNPWNY 399
|||||
QY 240 GESSLSVSKWNTFLKAMLVCSDAATNFRNLQDVFLLPDPGQWRDTRVYGVFSNPWNY 299
Db 330 SAVCVYSLGDIKVFRTSSLGKHYHSSLPNPRGKCLPDQOPIPTETFOVADRHPEVAQRV 399
|||||
QY 300 SAVCVYSLGDIKVFRTSSLGKHYHSSLPNPRGKCLPDQOPIPTETFOVADRHPEVAQRV 359
Db 390 EPMG 393
```

```
QY 360 EPMG 363
|||||

RESULT 3
ID O64906 PRELIMINARY; PRT; 653 AA.
AC O64906;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALCAPAPHINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RX MEDLINE; 97201573.
RA ENSER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RX MEDLINE; 97404659.
RA ENSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RA ENSER A., PFLANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U18243; G1000717; -.
DR EMBL; AF005370; G2337970; -.
SQ SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match 43.8%; Score 1865; DB 14; Length 653;
Best Local Similarity 46.7%; Pred. No. 0.00e+00;
Matches 270; Conservative 104; Mismatches 191; Indels 13; Gaps 13;

Db 57 LLMIL-SAITAASRFRIDKPRILVNLTDGFGQHRF-FGQPEHTVLFHLSNDSVYVGN 114
|||||
QY 1 LLLLLWAAASQAQHLRSRGPRIFAVWKGHVQDRVDFGQTEPTVLFHEPGSSSVWVGR 60
Db 115 NTIYLFDFAHSSNASTALNITSTHNTHRLSTCENFILLHNOTDGLACGTNSOKPSC 174
|||||
QY 61 GKVYLFDFPECKNASRTVNTVIGSTKGCQKQDCGNTITLLRRSEGLLACGTNARKPSC 120
Db 175 W-LINLLTTFGLPKLGLAPFSPSSGNLVLFQDNDTYSTINLYKSLSGS-HKFRRIAGOV 232
|||||
QY 121 WNLVNGTVPV-IGEMRGYAFSPDENSLVLFEGDEVYSTIRK-QEYNGKIPFRFRIGES 178
Db 233 ELYTSDTAMHRPQFVOATAVHKNESYDDKIYFFEQENSHSDFKQEPHTVPRYGVQVCSQ 292
|||||
QY 179 ELYTSDTVMQNPQFIKATIVHQDAYDDKIYFFREDNPKNPAPLNSVVAQLCRGDQ 238
Db 293 GESSLSVSKWNTFLKARLACVDYDTGRIYNELQDIFWQAPENSWEETLILGLFLSPWN 352
|||||
QY 239 GESSLSVSKWNTFLKAMLVCSDAATNFRNLQDVFLLPDPGQWRDTRVYGVFSNPWN 298
Db 353 FSAVCVETVKDIDHVFKTSKLNKHYHKLTPRPGCMKMHQHVPTTFQVADRYPEVADP 412
|||||
QY 299 YSAVCYSLGDIKVFRTSSLGKHYHSSLPNPRGKCLPDQOPIPTETFOVADRHPEVAQR 358
Db 413 VTQKNNAPEPIIQSKYIYTKLVYRVEYG-GVFWATIFYLTITKGTIHYIYVEDSNST 471
|||||
QY 359 VEPMGLKTPLFHSKYHYQKVAVHRMQASHGETFHVLYLTDDRGTHHKVVEPGEQEHSE 417
Db 472 ANILBINFQKAPATQIONILLNTNKLKYNSEWESEVPLDLCVYGNDCSCFMSRDP 531
|||||
QY 418 AFNIMEIOPFRRAAATQTHSLDAERKRLYVSSQWVEYSQVPLDLCVEYGGCHGCLMSRDP 477
```



```
Db 288 ARLVCSVPGEIGETHFDELQDVV-QOTQDV-RNPVIYAVFTSGSVFRGSVAVCYISMA 345
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 255 AMLVCS---DAATKNFNRLQDVLLPDPGQWRDRVYGVFSNP-WNY--SAVCVYSLG 308
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 346 DIRMVFNFGFAHKEGNYQWMPFSGKMPYRPGTCGPGTFTFSKSTKDYDPEVINFMR 405
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 309 DIKVFRTS-SLK-G--YH---SS-LPNRPGKCLPDQ-QP-I-PTETF--QVAD--R- 351
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 406 HPLMQAVYPLQRRPLVVRT-GAPYRLTTIAVDVDSADG-RYEVLFGLTDRGTQKVIV 463
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 352 HPEVAQRVEPMG--PLKTPLFHSKYHQYAVHMQASHGETFHVLYLTTDRGTHKVVE 409
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 464 LPKDDOEMELMLEEVEFKDPAPYKMTISSKROOLYVASAVGVTHLSLHRCQAYGAAC 523
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 410 -PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERKLYVSSQWEYSQVPLDLCEYGGC 468
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 524 ADCCLARDPYCAWDGQACSRYSASSKRRRRQDVHGNPIROCRFNSNANKNAVESQY 583
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 469 HGCLMSRDPYCGWDGRCIS-IYSSE-SVLQSNINPAEPHKECP--NPKPKAPLQKVS 524
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 584 GVAGSAAFLEQCPRSPQATVKW 605
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 525 A-PNSRYLSCPMESRHATYSW 545
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |

RESULT 11
ID O88633 PRELIMINARY; PRT: 754 AA.
AC O88633;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN IV ISOFORM A.
OS MUS MUSCULUS (MOUSE)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ECKHARDT F., MEYERHANS A.;
RT "Molecular cloning and expression pattern of a murine semaphorin
RT homologous to H-sema IV."
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF080091; G337768; -.
SQ SEQUENCE 754 AA; 85133 MW; EC05BDF3 CRC32;

Query Match 16.0%; Score 682; DB 11; Length 754;
Best Local Similarity 31.8%; Pred. No. 7.12e-141;
Matches 141; Conservative 119; Mismatches 140; Indels 43; Gaps 31;

Db 171 PYDPKLDTASALINEELIYAGVYI-DFMGTDAAIFRTLQKOTAMRTDQYNSRWLNPDSETH 229
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 139 PFPDENSLVLFEGDEVYSTIRKQYNGK-IPFRRIIRGESELYTS--DTV-MQNPQFIK 194
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 230 ABLIPDSAEKNDKLYFFFRSARAE-APONPAVYARIGRICLNDGGHCL-VNKNWSTFL 287
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 195 AIVHQD-QAYDDKIYFFREDNPKNPEAPLNVSRAQLCRDQGGESSLSVSKWTF 253
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 288 KARLVCSVPGEIGETHFDELQDVV-QOTQDI-RNPVIYAVFTSGSVFRGSVAVCYISM 345
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 254 KAMLVCS---DAATKNFNRLQDVLLPDPGQWRDRVYGVFSNP-WNY--SAVCVYSL 307
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 346 DIRMVFNFGFAHKEGPNYQWMPFSGKMPYRPGTCGPGTFTFSKSTKDYDPEVINFMR 405
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 308 GIDKVFRTS-SLK-G--YH---SS-LPNRPGKCLPDQ-QP-I-PTETF--QVAD--R- 351
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 406 THPLMQAVYPLQRRPLVVRT-GAPYRLTTIAVDVDSADG-RYEVLFGLTDRGTQKVIV 463
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 352 -HPEVAQRVEPMG--PLKTPLFHSKYHQYAVHMQASHGETFHVLYLTTDRGTHKV 408
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 464 VLPKDOOEVEELMLEEVEFKDPAPYKMTISSKROOLYVASAVGVTHLSLHRCQAYGA 523
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 409 E-PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERKLYVSSQWEYSQVPLDLCEYGG 467
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 524 CADCCCLARDPYCAWDGQACSRYSASSKRRRRQDVHGNPIRQCRGFSNANKNAVESQY 583
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
```

```
Qy 468 CHGCLMSRDPYCGWDGRCIS-IYSSE-SVLQSNINPAEPHKECP--NPKPKAPLQKVS 523
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 584 YGVAGSAAFLEQCPRSPQATVKW 506
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 524 LA-PNSRYLSCPMESRHATYSW 545
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |

RESULT 12
ID Q90607 PRELIMINARY; PRT: 772 AA.
AC Q90607;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COLLAPSEIN.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE; 94006554.
RA LUO Y., RAIBLE D., RAPER J.A.;
RT "Collapsein: a protein in brain that induces the collapse and
RT paralyis of neuronal growth cones."
RL CELL 75:217-227(1993).
DR EMBL; U02528; G410079; -.
DR PFAM; PF00047; 19; 1.
SQ SEQUENCE 772 AA; 88867 MW; BC6CB9E8 CRC32;

Query Match 16.0%; Score 681; DB 13; Length 772;
Best Local Similarity 31.5%; Pred. No. 1.31e-140;
Matches 146; Conservative 119; Mismatches 150; Indels 48; Gaps 38;

Db 166 RKSPYDPKLTATSLVDPGLYSYGA-ADFMGRDFAIFRTLGHHPHPIRTEQHDRLNDP 224
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 135 RGYAPFSPDENSLVLFEGDEVYSTIRKQYNGK-IPFRRIIRGESELYTS--DTV-MQNP 190
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 225 RFISAHLIPESDNPEDDKIYFFFRENAIDGHTGATARIQICIKNDGEGHRS-L-VNWK 283
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 191 QFIKATIVHQ-DQAYDDKIYFFREDNPKNPEAPLNVSRAQLCRDQGGESSLSVSKW 249
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 284 TFLKARLICVPGNGIDTHFDELQDVFLM-NSKDP-KNPVIYGVFTTSSNIFGSVAVC 341
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 250 NTFKLAMVCS--DA-ATKNFNRLQDVLLPDPGQWRDRVYGVFSNPW-N--SVC 303
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 342 MYSMTDVRRLVGLPYAHRDGNPYQWYQVYGRVYPRPGTCPSKTFGGEDSTKDLDPDEVIT 401
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 304 VYSLGIDKVFRTS-SLK-G--YH--S--S-LPNRPGKCLPDQ-QPI-PTETF--QV-A 349
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 402 FARSHPAMYNVPFINS-RPIMIKTDVDYQFQIVVDRVDAEDGQ-YDVMFTGTDIGTVL 459
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 350 -DR-HPEVAQRVEPMGKPLTFLFSK--YHQYKVAVHMQASHGETFHVLYLTTDRGTH 405
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 460 KVVSPKPTWLELEVLLEEMTVFREPVISAMKISTKQOOLYIGSATGVSQPLHRCVDV 519
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 406 KVEE-PGQEHSAFNINE-IQPFRAAAIQTMSLDAERKLYVSSQWEYSQVPLDLCEV 463
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 520 YGKACAECLARDPYCAWDGSCSRYSRYFTAKRRTRRQIRNGDPLTHCSDLQHHNPSQ 579
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 464 YGGGCHGLMSRDPYCGWDGRCISYSS-RSVL-QSINPAEPHKECPN-PKPKAP-- 518
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 580 TLEEKIIYGVNSSTFLFECSPKSQRAIYV-WQFQKQNDHDKVE 621
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQSC 556
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |

RESULT 13
ID O88632 PRELIMINARY; PRT: 785 AA.
AC O88632;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
```



```
Qy 139 PFSPEDSLVLFEDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MONPOFIK 194
Db 261 AELIPDSAEERNDKLYFEFFERSAE-APOSNAVYARIGRICLNDGCGHCL-VNKNWSTEL 318
Qy 195 ATIVHQD-QAYDDKIYFFREDNPKPEAPLVNVRVAQLCRGQGGESSLSVSKWNTL 253
Db 319 KARLVCSVPGEDGIETHFDELQDFV-QQOTDV-RNPVIYAVFTSSGVSFSGSACVYISM 376
Qy 254 KAMLVCS---DAATNKNENRLQDFVLLPDPSGQWRDTRVYGVFSNP-WNY--SACVCYSL 307
Db 377 ADIRWVNGPFAHKEGPNYQWMPFSGKMPYPRPCTCPGGTFTPSMKSTKDYDPDEVINMR 436
Qy 308 GDIDKVFRTS-SLK-G--YH-----SS-LPNRPGKCLPDQ-QP-I-PTETF--QVAD--R 351
Db 437 SHPLMYQAVYPLORRPLVVRT-GAPYRLTTIADVDOVDAGDG-RYEVLFLGTDRGTQKVI 494
Qy 352 -HPEVAORVEPMG--PLKTPLEHSHYQKVAVHRMQASHGETFHVLYLTDRGTIHKV 408
Db 495 VLPKDDQEMEELMEEVEVFKDPAPVKMTITSSKRQOLYVASAVGVTHLSLHRCQAYGAA 554
Qy 409 E-PGEQEHSEFAFNIMEIQPFRAAAIQTMSLDAERKLYSSQWEVSQVPLDLCEVYGGG 467
Db 555 CADCCCLARDPYCAWDGQACSRYTASSKRRSRRQDVHRGNPIRQCRGFNSNANKNAVESVQ 614
Qy 468 CHGCLMSRDPYCGWDQGRCS- IYSSER-SVLQSIINPAEPHKECP--NPKPKAPLQKVS 523
Db 615 YGVAGSAAFLECPQSPQATVKW 637
Qy 524 LA-PNSRYLLSCPMESRHATYSW 545
```

Search completed: Thu Jul 8 19:42:28 1999
Job time : 47 secs.

WQESREH

(TW)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:47:28 1999; MasPar time 29.15 Seconds
Tabular output not generated.
412.999 Million cell updates/sec

Title: >US-09-041-236-2
Perfect Score: 4211 from US09041236.pep (40 of 45)
Sequence: 1 LLLLLWAAASAGHLRSGP.....HKENVQSCPEGHQSPNCIL 566
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.230; Variance 148.600; scale 0.244

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	656	15.6	771	13	R71380 Human semaphorin III	4.04e-52
2	615	14.6	477	13	R74175 Human collapsin.	5.37e-48
3	612	14.5	775	33	W63748 Human semaphorin.	1.07e-47
4	606	14.4	861	22	W17658 Mouse CD100 antigen.	4.29e-47
5	599	14.2	861	32	W58540 Human semaphorin.	2.16e-46
6	544	12.9	776	32	W51313 Human semaphorin W.	6.81e-41
7	538	12.8	862	22	W17657 Human CD100 antigen.	2.70e-40
8	463	11.0	441	13	R71381 Vaccinia virus semaph	7.35e-33
9	459	10.9	730	13	R71379 Grasshopper semaphori	1.82e-32
10	428	10.2	587	32	W51314 Human semaphorin W.	2.03e-29
11	390	9.3	974	33	W64221 Human secreted protei	1.04e-25
12	378	9.0	712	13	R71384 Tribolium semaphorin	1.53e-24
13	367	8.7	650	13	R71382 Drosophila semaphorin	1.77e-23
14	362	8.6	724	13	R71383 Drosophila semaphorin	5.40e-23
15	352	8.4	930	32	W57260 Human semaphorin Y.	4.98e-22
16	325	7.7	888	25	W19857 Human semaphorin Z.	1.94e-19

17	316	7.5	887	25	W19856 Rat semaphorin Z.	1.40e-18
18	296	7.0	929	32	W57259 Rat semaphorin Y.	1.11e-16
19	157	3.7	122	13	R71385 Variola major virus s	4.50e-04
20	107	2.5	111	32	W51315 Human semaphorin W pa	4.93e+00
21	99	2.4	591	4	R23006 Protein transcribed f	1.95e+01
22	101	2.4	2329	25	W25038 Partial BRCA2 cancer	1.39e+01
23	98	2.3	832	33	W61092 Taq DNA polymerase I	2.30e+01
24	96	2.3	832	33	W61090 Taq DNA polymerase I	3.22e+01
25	91	2.2	478	1	R04881 Recombinant elastase.	7.33e+01
26	91	2.2	598	9	R48631 Sequence of nuclear r	7.33e+01
27	92	2.2	832	33	W61091 Taq DNA polymerase I	6.23e+01
28	92	2.2	854	30	W56309 Clas II S-receptor ki	6.23e+01
29	92	2.2	855	6	R29815 S. receptor kinase pro	6.23e+01
30	93	2.2	4572	30	W2845 A. mediterranei rifam	5.29e+01
31	89	2.1	534	25	W25031 Partial BRCA2 cancer	1.01e+02
32	89	2.1	554	13	R66209 Novel thermostable DN	1.01e+02
33	89	2.1	638	5	R03924 E. coli HSP (dnaK).	8.63e+01
34	89	2.1	680	4	R23143 Mutant thermostable D	1.01e+02
35	89	2.1	832	13	R76693 DNA-polymerase (F73L,	1.01e+02
36	89	2.1	832	1	P90556 Purified native therm	1.01e+02
37	89	2.1	832	33	W61087 Taq DNA polymerase I	1.01e+02
38	89	2.1	832	33	W61088 Taq DNA-polymerase I	1.01e+02
39	89	2.1	832	13	R76690 Taq DNA-polymerase RE	1.01e+02
40	89	2.1	832	33	W61089 Taq DNA polymerase I	1.01e+02
41	89	2.1	833	36	W5942 Amino acid sequence o	1.01e+02
42	89	2.1	833	27	W24211 Cleavase DN nuclease.	1.01e+02
43	89	2.1	833	27	W24212 Cleavase DA nuclease.	1.01e+02
44	89	2.1	3418	26	W19211 Human breast cancer s	1.01e+02
45	89	2.1	3418	26	W23287 Human breast and ovar	1.01e+02

ALIGNMENTS

RESULT 1

ID R71380 standard; Protein; 771 AA.
AC R71380;
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN WO9507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 60-63; 101pp; English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA;

Query Match 15.6%; Score 656; DB 13; Length 771;
Best Local Similarity 31.2%; Pred. No. 4.04e-52;
Matches 143; Conservative 115; Mismatches 155; Indels 45; Gaps 35;

Db 166 rgkspdydkltasllldgelysgta-adfmgdrdfairtghhhpirtqdhsrwlndp 224

QY 29 HVGDRVD-FQGTPEHTVLFH--EPGSSVWVGGRGVLYLDFEPGKN-ASVRTV-NI-G 82
 Db 92 ahrelcykgsedcnyivrlakidddrvlicgtnaykplcrhyalkdgyvvekeye 151
 QY 83 STKSC-L-DKR--DCENYITLLRERREG-LLACGTWNRHPSWN--LVNGT-VVPLG-E 133
 Db 152 grgicpdpdhnstalysegqlsyatv-adfsqtdp-li-yrg-p-lrteradi-k-qln 204
 QY 134 MRGVAPSPDENSLVLFEGDEYVSTIRKOEYNGKIPREFRRIRGESELYTSDTVWQNPQFI 193
 Db 205 apnfvn-tmeyndfiffretaveyincgkaly-srvarvckhdkgg-phqggdrwtstf 261
 QY 194 KATTVHQDQYDDKXYFFREDNPD-KNPEAPLNVSRAQLCRDQGGESSLSVSKWTF 252
 Db 262 lkrlncsvpdgyfyfneqlgstdiiegnvgq-vekllygvfttpvnsiggsavcafs 320
 QY 253 LKAMLVCSDAATNK-NFNRLQDVF-LUPDP--SGQWRDTRVYGVFSNPNW-Y--SAVCVYS 306
 Db 321 mkstlesfdgpfkeqetmnsnlavpslkvppeprpgqcvndsrtdlpdvsnfvkshltmd 380
 QY 307 L-G--D-IDKVFRT--SSLKG-YHS--SL--PNRPGKCLPDQOPIPTETFOVADRHPVEA 356
 Db 381 eavpafft-rpallrlisqlyftfiavdqvrtpdgdqaydvlfigtddgkvikalnsasf 439
 QY 357 QRVEPMGLKTPLFH-S-KYHQKAV-HRMQASHGETFHYLYLTDRGTIHKYVEPG-- 411
 Db 440 dssdtdsvleelqvlppgvpknlyvrmddgdsdklvvsvddeiilaiklrcgskit 499
 QY 412 EQEHSFAFNIE-IQPFRAAAIOTM--SLDAERRKLYVSSQWESQVPLDLCEVYG-G 466
 Db 500 ncrevslqdpycawdnvelkctavgsdpws 530
 QY 467 GCHCLMSRDPYCGWDQ-G-RCIIYSERS 495

RESULT 10
 ID W51314 standard; Protein; 587 AA.
 AC W51314;
 DT 08-SEP-1998 (first entry)
 DE Human semaphorin W.
 KW Human; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
 KW immunosuppressant; gene therapy; diagnosis; research reagent.
 OS Homo sapiens.
 PN WO9815628-A1.
 PD 16-APR-1998.
 PF 03-OCT-1997; J03549.
 PR 09-OCT-1996; JP-287636.
 PA (SUMO) SUMITOMO PHARM CO LTD.
 PI Kikuchi K, Kimura T;
 DR WPI: 98-261015/23.
 DR N-PSDB: V07281, V07282.
 PT Nerve extension inhibitor protein semaphorin W - is useful as
 PT therapeutic drug and diagnostic and research reagent
 PS Example 4; Page 69-72; 90pp; Japanese.
 CC The present sequence represents human semaphorin W. Semaphorin W and
 CC its derivatives are nerve extension inhibitors which are useful as
 CC antiallergic, immunosuppressant and anticancer agents. The DNA
 CC encoding semaphorin W can also be used in gene therapy, e.g. using
 CC a viral vector. The proteins, peptides, DNA and antibodies which
 CC recognise the protein or peptides, can be used as diagnostic or
 CC research reagents. Semaphorin W can be used as a screen for
 CC semaphorin W antagonists with possible therapeutic use.
 SQ Sequence 587 AA;

Query Match 10.2%; Score 428; DB 32; Length 587;
 Best Local Similarity 28.9%; Pred. No. 2.03e-29;
 Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgddeiyffittetsrafsyerikvprvarcagdlgrktlq-grwtflkadiilcp 115
 QY 201 DQAYDDKIYFFREDNPDKNPEAPLNVSRAQLCRDQGGESSLSVSKWTFKAMLVCS 260
 Db 116 gpehgrassvlqdvavirpelga-gtbfygfssqvegatisavcafrpqdirtvngp 174

QY 261 DAATNKNFNRLQDVFLLPDPSPGQWRDTRVYGVFSNPNW--YSAVCVYSLGDIKVERTS 317
 Db 175 frelkhdncrglpvwdndvpqprgpcitnmklrhfgsslsipdrvlttfirdhplndrp 234
 QY 318 --SLK-GYHSLP--N---PRGKCLPDQOPIP--TETFOVADRHP-E-V-AQR-VE-P 361
 Db 235 vfpadgphllvtttdtaylrvvahrvtslsgkeydvlylgtedghlhravriga-ls-vl 292
 QY 362 MGPLKT-PLFHSK-YHQKAVAHMQASHGETFHYLYLTDRGTIHKVVEPGEHSEFAF 419
 Db 293 edial--fpeqpqvemkml-yhsw-llvgsrtevtqvtntncgrlqs-csecillaqdpvc 347
 QY 420 NIMEIQPFRAAAIOTMSLDAERRKLYVSSQWESQVPLDLCEVYGGCHGLMSRDPYC 479
 Db 348 awafriidccvahahegrhlvgdiesadvslcpkegprpvpfvvataahvvlpcps 407
 QY 480 GWD-Q-GRCISIIYSERSVLSQINPAEPHKECPNPKAPLOKVS LAPNSRYLSCPME 537
 Db 408 sawascvwhqpsv 421
 QY 538 SRHATYSRRHKENV 551

RESULT 11
 ID W64221 standard; Protein; 974 AA.
 AC W64221;
 DT 06-OCT-1998 (first entry)
 DE Human secreted protein from clone CJ145_1.
 KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
 KW cell proliferation; differentiation; immune system; suppressor; ligand;
 KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
 KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
 KW anti-inflammatory.
 OS Homo sapiens.
 PN WO9827205-A2.
 PD 25-JUN-1998.
 PF 17-DEC-1997; U23330.
 PR 16-DEC-1997; US-991872.
 PR 18-DEC-1996; US-769192.
 PR 13-JAN-1997; US-783401.
 PA (GEM) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI: 98-362774/31.
 DR N-PSDB: V44295.
 PT New polynucleotides and secreted proteins - obtained from human
 PT foetal brain, human adult testes, human adult brain and human adult
 PT salivary gland cDNA libraries
 PS Claim 17; Page 71-74; 110pp; English.
 CC This sequence represents a novel secreted protein from clone CJ145_1
 CC isolated from a human fetal brain cDNA library. This protein has
 CC applications for nutritional use, cytokine and cell
 CC proliferation/differentiation activity, immune stimulating or
 CC suppressing activity, hematopoiesis regulating activity, tissue growth
 CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
 CC haemostatic and thrombotic activity, receptor/ligand activity,
 CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
 CC tumour inhibition activity and other activities.
 SQ Sequence 974 AA;

Query Match 9.3%; Score 390; DB 33; Length 974;
 Best Local Similarity 29.8%; Pred. No. 1.04e-25;
 Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechmfikvllknddalfvcgtfnafnpscrnykmdtlfegdfsgmarcpdykhan 172
 QY 91 KRDCENVIT-LLERRSEGLACCTNARHPSCWNLVNGTVVPLG-EMRGYA--PSPSPENS 146
 Db 173 valfadgklysatvt-dflaidaviyrs-l-gesp--tlrtvkhdkskwlkpfyqvadvy 228
 QY 147 LVLFEGDEVYSTIRKQBYNG-KIPFRFRIRGESELYTSDTVWQNPQFIKATIVHQDQAYD 205

OY 318 -SLKG-Y-H-S-LPNRPGKCLP-DQOPIPTETFOVADRHPE-V-AQRV-EP-MGPLK 366
Db 411 hqpllttsralltqvavdgmagphsn-itvmflgsndgtvkvltpggrsggpepille 469
OY 367 -TPLFH--SKYHYQKVAVHRMQASHGETFHVLYLTTRGTIHKVVEPGEQHSFAFNIME 423
Db 470 eiddaysparcsgkrtagtarriigleldteghrlfvafsgcivylplsracarhga-cgrs 528
OY 424 -IQPF---R---RAA--A--IOTMSLDAERRKLYVSSQWEVSQVPLDLCEVYGGCH-G 470
Db 529 clasqdpycgwhsrgcvidlrgsggtddvqagngesmehgdcqd 572
OY 471 CLMSRDPYCGWDOGR-CISIYSSE-RSVLQSNPAE-PHKECPN 511

Search completed: Thu Jul 8 19:48:05 1999
Job time : 37 secs.

THIS PAGE BLANK (USPTO)

WORLD (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:46:39 1999; MasPar time 25.88 Seconds
Tabular output not generated. 876.232 Million cell updates/sec

Title: >US-09-041-236-2
Description: (1-566) from US09041236.pap (40 of 45)
Perfect Score: 4211
Sequence: 1 LLLLLWAAAASAOQHLSRGP.....HKENVQSCPEGHQSPNCIL 566

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 48.941; Variance 86.790; scale 0.564

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	1850	43.9	553	2 T03102 semaphorin homolog A3	0.00e+00
2	737	17.5	748	2 I48744 semaphorin A - mouse	6.57e-134
3	725	17.2	751	2 I48748 semaphorin E - mouse	3.69e-131
4	716	17.0	749	2 G01856 semaphorin V - human	4.24e-129
5	688	16.3	753	2 G02173 semaphorin III family	1.06e-122
6	681	16.2	772	2 A48069 collapsein - chicken	4.19e-121
7	669	15.9	772	2 I48747 semaphorin D - mouse	2.27e-118
8	662	15.7	666	2 I58169 semaphorin III - mouse	8.89e-117
9	656	15.6	771	2 D49423 semaphorin III precursor	2.06e-115
10	618	14.7	782	2 I48746 semaphorin C - mouse	8.56e-107
11	522	12.4	834	2 S66498 M-sema F protein precursor	3.16e-85
12	474	11.3	760	2 I48745 semaphorin B - mouse	1.40e-74
13	470	11.2	403	2 E42521 A39R protein - vaccin	1.06e-73
14	463	11.0	441	2 S29921 fasciclin IV precursor	3.70e-72
15	461	10.9	730	2 JH0798 semaphorin F precursor	1.02e-71
16	448	10.6	1074	2 JCS528 semaphorin I precursor	7.30e-69
17	389	8.2	711	2 A49423 semaphorin I - fruit	4.89e-56
18	367	8.7	656	2 C49423 semaphorin II precursor	2.56e-51
19	362	8.6	724	2 C49423 Sall9R protein - vacc	2.99e-50
20	293	7.0	295	2 JQ1775 plexin 2 precursor -	8.82e-36
21	196	4.7	1884	2 JCI4975 plexin - African claw	1.26e-16
22	171	4.1	1905	2 I51553 plexin 3 precursor -	4.63e-12
23	170	4.0	1872	2 JCI4976 plexin 3 precursor -	6.97e-12

24	168	4.0	1894	2 JQ4980 plexin 1 precursor -	1.58e-11
25	157	3.7	122	2 JQ1845 I4R protein - variola	1.32e-09
26	157	3.7	122	2 H36852 A43R protein - variol	1.32e-09
27	155	3.7	142	2 JQ1776 SallR protein - vacc	2.92e-09
28	120	2.8	775	2 E70320 polyribonucleotide nu	1.37e-03
29	105	2.5	227	2 S09922 hypothetical protein	2.03e-01
30	105	2.5	235	1 Q0BEC9 HXLF4 protein - human	2.03e-01
31	102	2.4	275	2 A34866 T-cell surface protei	5.20e-01
32	99	2.4	275	2 S08464 T-cell alloantigen RT	1.31e+00
33	103	2.4	406	2 B69064 conserved hypotherm	3.81e-01
34	99	2.4	460	2 S35772 translation elongatio	1.31e+00
35	99	2.4	563	2 S78224 virulence-associated	1.31e+00
36	99	2.4	591	2 S26565 virulence-associated	1.31e+00
37	99	2.4	591	2 S09498 virulence-associated	1.31e+00
38	99	2.4	591	2 S28664 virulence-associated	1.31e+00
39	99	2.4	593	2 S15215 virulence-associated	1.31e+00
40	99	2.4	595	2 A46050 thyroid/steroid recep	1.31e+00
41	99	2.4	1375	2 JCS148 hepatocyte growth fac	1.31e+00
42	98	2.3	641	2 S32017 flagellum-associated	1.77e+00
43	98	2.3	970	2 C57282 ankyrin-related prote	1.77e+00
44	98	2.3	979	2 B57282 ankyrin-related prote	1.77e+00
45	98	2.3	1786	2 A57282 ankyrin-related prote	1.77e+00

ALIGNMENTS

RESULT	1
ENTRY	T03102
TITLE	semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM	#formal_name alcelaphine herpesvirus 1
DATE	24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS	T03102
REFERENCE	214840
#authors	Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal	J. Virol. (1997) 71:6517-6525
#title	Primary structure of the alcelaphine herpesvirus 1 genome.
#accession	T03102
#status	preliminary; translated from GB/EMBL/DBJ
#molecule_type	DNA
#residues	1-653
#label ENS	1-653
#cross-references	EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY	#length 653 #molecular-weight 73645 #checksum 5501
Query Match	43.9%; Score 1850; DB 2; Length 653;
Best Local Similarity	46.8%; Pred. No. 0.00e+00;
Matches	267; Conservative 102; Mismatches 189; Indels 13; Gaps 13;

Db	57	LLMIL-SAITAAKSRFDKPLIYNLTDGFGQHRF-FGQEPHTVLFHSLNSDYVYCGN	114
Qy	1	LLLLLWAAAASAOQHLSRGPRIFAVWKGHVQDQVDFGQTEPHTVLFHEPGSSVYVWVGR	60
Db	115	NTIYLFDAHSSNASTALINITSTHRLSSTCENFTLLHNOTDGLLACTNSQKPS	174
Qy	61	GKVLDFPEPEKKNASRVTVNGSTGCLDRDCENYTLERSEGLLACTNARHPS	120
Db	175	W-LINNLTFQPLKGLAPFSSGNLVLFDQNDTYSTINLYKSLGS-HKFRRIAGQV	232
Qy	121	WNLVNGTVVP-LGEMRGVAPSPDENSLVLFEGDEVYSTIRK-QEYNGKIPRFRIGES	178
Db	233	ELYTSDTAMHRPQFOQATAVHKNSYDDKIYFFQENSHSDFKOPHTVPRVQVCS	292
Qy	179	ELYTSDTVMQNPQIKATIVHQDQAYDDKIYFFREDNPDKNPEAPLNVSVQALCR	238
Db	293	GGESLSVYKWTTLKALACVDYDTGRIYNELODIFTWQAPENSWEETLYLGLFSP	352
Qy	239	GGESLSVSKWNTTLKAMLVCSDAATNKNFRLQDVLDPDPGQWRDTRVYGFSP	298
Db	353	FSAYCVETVKDIDHVFSTSKLKNYHHKLPTPRPGQCKMKNHQHVTETFOVADRY	412
Qy	299	YSACVYSLGDIKVFRTSSKLGYSLLPNRPCKLDPDQPIPTETFOVADRYEVA	358
Db	413	VYQNNAMFPIQSKIYIVTKLLVRYVEYG-GVFWATIFYLTITKGTIHYVYEDSN	471

QY 359 VEPGPKLPLFHSKYHYQKVAVRMOASHGETFH-VLYLTDRGTHIKVVEPGEQHSF 417
Db 472 ALNTEINPKQKPAIONILLNTNLKLYNSEWESEVPLDLCXSVYVGNDCFCSEMSRDP 531
QY 418 AFNIMEIQPFRAAAIQTMSLDAERRLKYVSSQWESQVPLDJCEVYGGCHGCLMSRDP 477
Db 532 LCTWYNNTC-S-FK-QRVSVETGPGNARTLSEMGDHYAPTVMKHQVSIPLLSNLSLSCP 588
QY 478 YCGWDGRCISYSSERSVLQSNPAEPH-KE-CPNPKPKAPLOKQVSLAPNSRYILSCP 535
Db 589 AVSNHADYFWTKOGFTEKRCVHTKNDICIL 619
QY 536 MESRHATYSWRHRENVPQSCPEGHQSPNCIL 566

RESULT 2 I48744 #type complete
ENTRY semaphorin A - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
DATE 04-Sep-1998

ACCESSIONS I48744
REFERENCE I48744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48744
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-748 #label RES
#cross-references EMBL:X85990; NID:g854323; PID:g854324

GENETICS
#gene sema
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 748 #molecular-weight 82894 #checksum 9017

Query Match 17.5%; Score 737; DB 2; Length 748;
Best Local Similarity 33.78; Pred. No. 6.57e-134;
Matches 168; Conservative 112; Mismatches 166; Indels 52; Gaps 36;

Db 112 ECMNFVRLHAYNHTLLACRTGAFHPTCALWRWATAGGTHASTGPEKLEDGKGTYPDP 171
QY 93 DCENYITLLRRSEGL-LLACGTNARHPSG---WNLVNGI-V-V-P--LGEMRGIAPFSP 142
Db 172 RHRPPSVLGEELYSGV-TADLMGRDFTIFRSLGQNPSTLTPHDSRWLNPEKFKVFWI 230
QY 143 DENSLVLFEGDEVYSTIRKQYNGK-IPFRFRRIRGESELYT-S-DTV-MQNPFQIKAT-I 197
Db 231 PESENPDDDKIYFFERESAEEAAPAMGRMSVRVGOICRNDLGGORSL-VNKKWTTFLKAR 289
QY 198 VHODQAVDDKIYFFEREDNDPKNPE-APLNSRVAQLCRDQGGESLSVSKWNTFLKAM 256
Db 290 LVCSPVQVEGDTHFDQLOQVFLLS-SR-DRQTPLLXAVFTSSGVFQGSVAVCYMSNDVR 347
QY 257 LVCS--DAATNKNFNRLQDVLLPDPGQWRDRTRYGVFSNPNW-Y--SAVCVYSLGDID 311
Db 348 RAFLGPLPHKEGTHQWVSQGRVPPRCMCPKSTFGTFSSKDPDPDDVVIQGRNHPLM 407
QY 312 KVFTS-SLK-G-----YHSSLNPRPGKCLPDQ-QPIP-TETF--QVAD--R-HPEV 355
Db 408 YNPVLPMSG-R-PLFLQVAGYFTFQIAADRAVAADGCH-YDVLFIGTDGTVLKVLSVPK 464
QY 356 AQRVEPMGPKLPLF-H--SKYHYQKVAVRMOASHGETFHVLYLTDRGTHIKVVE-P- 410
Db 465 GRPNSEGLLEELQVFEDESAITSMQISSKROOLYVYASRAAQAIALHRCTALGRACAE 524
QY 411 GEGEHSFAFNIMEIQPFRAAAIQTMSLDAERRLKYVSSQWESQVPLDLCVYGGCHG 470
Db 525 CCLARDPYCAWDSACTRFOPTAKRRFRRODIRNGDPSTLCSGDSHSLVLEKVKLVGES 584

QY 471 CLMSRDPYCGWDGRCISYSSERSVL--QSNPAEPHKECPNPKPKAPLOKQVSLAPNS 528
Db 585 GSAFLECEPRSLQAHVOW 602
QY 529 -RYLSCPMEHRATYSW 545

RESULT 3 I48748 #type complete
ENTRY semaphorin E - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
DATE 04-Sep-1998

ACCESSIONS I48748
REFERENCE I48748
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48748
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-751 #label RES
#cross-references EMBL:X85994; NID:g854331; PID:g854332

GENETICS
#gene sema
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 751 #molecular-weight 85259 #checksum 8961

Query Match 17.2%; Score 725; DB 2; Length 751;
Best Local Similarity 33.0%; Pred. No. 3.69e-131;
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;

Db 167 SNNPNVTVSVMINBELFSGMYI-DFMGTDAAIFRSLTKRMQLRDTDOHNSKWLSEPMFVD 225
QY 139 PFSPDENSLVLFEGDEVYSTIRKQYNGK-IPFRFRRIRGESELYTS--DTV-MQNPFQIK 194
Db 226 AHVIPDGTDPNDAKVYFFEKRLTDNNRSTKQIHSMIARICPNNDTGGORSL-VNKKWTTFL 284
QY 195 ATIVHOD-QAYDDKIYFFEREDNDPKNPEAPLNSRVAOLCRDQGGESLSVSKWNTFL 253
Db 285 KARLVCSVTDEGPETHFDELEDVLL-ETDNP-RITLYVIGITTSVSVFKGSVAVCYHL 342
QY 254 KAMLVCS--DA-ATNKNFNRLQDVLLPDPGQWRDRTRYGVFSNPNW-Y--SAVCVYSL 307
Db 343 SDIQVFNFGPFAHKEGPNHQLISYQGRIPYPRGTCPGGAFTPNMRTTKDFDDVVFTIR 402
QY 308 GDIDKVFRTS-SLK-G-----YHSSLNPRPGKCLPDQ-QP-I-PTETF--QVADRH 353
Db 403 NHPLMYSISPITHRRPLIVRIGTDKYTKIAYDRVNAADG-RYHVLFLGTDGTVOKVVV 461
QY 354 EVAQRVEPMGPKL-KTPLF-H--SKYHYQKVAVRMOASHGETFHVLYLTDRGTHIKVVE 409
Db 462 LPTNSASGELLELEVEFNHVPITIMEISSKKOOLYVSSNEGVSQVSLHRCHRYGTAC 521
QY 410 -PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERRLKYVSSQWESQVPLDLCVYGGGC 468
Db 522 ADCCLARDPYCAWDSCHSRFPTGKRSSRRQDRVHGNPLTQCRGFNLKAYRNAEIVQY 581
QY 469 HGCLMSRDPYCGWDGRCISYSS--RSVLOSINPAEPHKECP--NPKPADPKLQKVL 524
Db 582 GYRNSTFLECAPKSPQASIKWLQKDK 609
QY 525 A-PNSRYILSCPMEHRATYSWR-HKEN 550

RESULT 4
ENTRY semaphorin V - human
TITLE #formal_name Homo sapiens #common_name man
ORGANISM 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
DATE 04-Sep-1998

[illegible]


```
ORGANISM      #formal_name vaccinia virus
DATE          20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
              09-Sep-1997
ACCESSIONS    S29921
REFERENCE      S29907
#authors      Amegadzie, B.Y.
#submission   submitted to the EMBL Data Library, January 1991
#accession    S29921
#status       preliminary
#molecule_type DNA
#residues     1-441 #label AME
#cross-references EMBL:X57318; NID:g62239; PID:g62254
SUMMARY       #length 441 #molecular-weight 50185 #checksum 6034

Query Match      11.0%; Score 463; DB 2; Length 441;
Best Local Similarity 33.9%; Pred. No. 3.70e-72;
Matches 87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;

Db 115 LVCGTNGNPKWK-IDGDDPKHGRGYAPYONSKVTIISHNGC-VLSDINISK-EG-I 170
QY 109 LAGCTNARHSCNWLNVGTVPVLEMGYAPFSPDENSLVLFEGDEVYSTIRKOEYNGKI 168

Db 171 KMRREFDPCGYDLTADNVIPKDG-LRGAFVDKGTID-KVILFTDTIGSKR--I-VK 225
QY 169 PRFRIRGE-S-ELYTSDTVMQNPQFIKATIVHODQAYDDKIYFFREDNPDKNPEAPLN 226

Db 226 IPYIAQCLNDEGSSLSHRNSTFKVLEEC-DID-GRSY-R-Q-IIHSRIKTD-ND 279
QY 227 VSRVAQLCRGDQGESLSVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLLPDPGQWRD 286

Db 280 TILVVFDDPYVSKALCTYSMTIKQSFSTKLEGYTKQLPSPASGICLPAGKVPVHTTF 339
QY 287 TRYGVFSNPWNSAVCVISLGDIDKVFRTSSUKGYHSSLPNRPKGKCLPDQPIPTET 346

Db 340 EVIEKYNVLDDIIKPLS 356
QY 347 QVADRHPEVAQVPEPMG 363

RESULT 15
ENTRY      JH0798      #type complete
TITLE      fasciclin IV precursor - American bird grasshopper
ORGANISM   #formal_name Schistocerca americana #common_name American
           bird grasshopper
DATE       30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
           09-Sep-1997
ACCESSIONS JH0798
REFERENCE   JH0798
#authors   Kolodkin, A.; Matthes, D.J.; O'Connor, T.P.; Patel, N.H.;
           Admon, A.; Bentley, D.; Goodman, C.S.
#journal   Neuron (1992) 9:831-845
#title     Fasciclin IV: Sequence, expression, and function during
           growth cone guidance in the grasshopper embryo.
#accession JH0798
#molecule_type mRNA
#residues  1-730 #label KOL
#cross-references GB:L00709; NID:g160844; PID:g160845
#experimental_source embryo
COMMENT    This protein plays a role in growth cone guidance in the developing
           central nervous system.
KEYWORDS   glycoprotein; transmembrane protein
FEATURE    1-22      #domain signal sequence #status predicted #label sig\
23-730      #product fasciclin IV #status predicted #label MAT\
23-627      #domain extracellular #status predicted #label EXT\
628-652     #domain transmembrane #status predicted #label TM\
653-730     #domain intracellular #status predicted #label INT\
44,71,163,267,360,
539         #binding_site carbohydrate (Asn) (covalent) #status
           predicted
SUMMARY    #length 730 #molecular-weight 81214 #checksum 5881

Query Match      10.9%; Score 461; DB 2; Length 730;
```

```
Best Local Similarity 28.0%; Pred. No. 1.02e-71;
Matches 143; Conservative 130; Mismatches 182; Indels 56; Gaps 46;

Db 32 QFGEIRVQREFLGNESHKDHFKLEKDHNSLLVGGARNIVYINISRLDTTEFTQRIEWHSSG 91
QY 29 HVGODRDV-FGQTEPHTVLFH--EPGSSSVWVGGRGVLYLDFDPEGKN-ASVRTV-NI-G 82

Db 92 AHRELCLYLGKSEDDCONYIRVLAKIDDDRLICGTNAYKPLCRHRYALKGDYVVEKEYE 151
QY 83 STRGSC-L-DKR--DCENYITLLRSEGL--LLACGTNARHPSCWN--LVNGT-VVPLG-E 133

Db 152 GRGLCPDPDHNSTAIYSEGLYSATV-ADFSGTDP-LI-YRG-P-LRTERSGL-K-QLN 204
QY 134 MRGYAFSPDENSLVLFEGDEVYSTIRKOEYNGKIPRRIRGESELYTSDTVMQNPQFI 193

Db 205 APNEVN-TMEYNDFIFFFFRETAVEYINCGKAIY-SRVARVCKHDKGGPHQFG-DRWTSF 261
QY 194 KATIVHODQAYDDKIYFFREDNPD-KNPEAPLNVSRAQLCRGQDQGESLSVSKWNTF 252

Db 262 LKSLNCSVPDYPFFYFNEIQTSDIIEGNYGGO-VEKLIYGVFTTTPVNSIGGSVAVCAFS 320
QY 253 LKAMLVCSDAATNK-NFNRLQDVF-LLPDP-SGQWRDTRVYGVFSNPWN-Y--SAVCVYS 306

Db 321 MKSILESEFDGPEKEQETMNSNWLAVPSLVKYPEPRPGOCVNDSTLDPVSYNFVKSHTLMD 380
QY 307 L-G--D-IDKVFT--SLKG-YHS--SL--PNRPKGKCLPDQPIPTETFOVADRHPEVA 356

Db 381 EAVPAFFT-RPILIRISLOYRFTKIAVDQOVTPDQKAYDVLFTGDDGKVIKALNSASF 439
QY 357 QRVEPMGKLTPLFH-S-KVHYQKVAV-HRMOASHGETFHVLYLTDRGTIHKVVEFG-- 411

Db 440 DSSDTVDVVVIELQVLPPGVPVKNLYVVRMDGDSKLVVVSDEILAIAIKLHRCGSKDIT 499
QY 412 EQEHSFAFNIME-IQPFRAAAIQT---SLDAERRKLYVSSQWESQVPLDLCEVYG-G 466

Db 500 NCRECYSLDOPYCAWDNVELKCTAVGSPDWS 530
QY 467 GCHGCLMSRDPYCGWDQ-G-RCISYSSERS 495
```

Search completed: Thu Jul 8 19:47:09 1999
Job time : 30 secs.

THIS PAGE BLANK (USPTO)

(TT)

Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	470	11.2	403	1	VA39_VACCC	PROTEIN A39.	4.48e-86
2	463	11.0	441	1	VA39_VACCC	PROTEIN A39.	2.81e-84
3	163	3.9	1871	1	SEX_HUMAN	TRANSMEMBRANE PROTEIN	9.40e-13
4	105	2.5	227	1	U08_HCMVA	HYPOTHETICAL PROTEIN H	1.75e-02
5	102	2.4	275	1	NT82_RAT	T-CELL ECTO-ADP-RIBOSY	3.10e-01
6	99	2.4	275	1	NT11_RAT	T-CELL ECTO-ADP-RIBOSY	3.12e-01
7	99	2.4	460	1	EF1A_TRIRE	ELONGATION FACTOR 1-AL	3.12e-01
8	99	2.4	591	1	VRP2_SALTY	65 KD VIRULENCE PROTEI	3.12e-01
9	99	2.4	591	1	VRP2_SALCH	65 KD VIRULENCE PROTEI	3.12e-01
10	99	2.4	591	1	VRP2_SALEN	65 KD VIRULENCE PROTEI	3.12e-01
11	99	2.4	593	1	VRP2_SALDU	65 KD VIRULENCE PROTEI	3.12e-01
12	99	2.4	916	1	PAW1_AJECB	PLASMA MEMBRANE ATPASE	3.12e-01
13	95	2.3	255	1	HA21_HUMAN	HLA CLASS II HISTOCOMP	1.21e+00
14	95	2.3	518	1	VL2_HPV5B	MINOR CAPSID PROTEIN L	1.21e+00
15	95	2.3	518	1	VL2_HPV05	MINOR CAPSID PROTEIN L	1.21e+00
16	95	2.3	683	1	ANOH_ARTGO	HISTAMINE OXIDASE (EC	1.21e+00
17	98	2.3	1132	1	DNBI_HSV6U	MAJOR DNA-BINDING PROT	4.40e-01
18	95	2.3	1390	1	MET_HUMAN	HEPATOCYTE GROWTH FACT	1.21e+00
19	94	2.2	329	1	R8SR_ECOLI	RIBOSE OPERON REPRESSO	1.68e+00
20	91	2.2	348	1	DDL_ENTFA	D-ALANINE--D-ALANINE L	4.44e+00
21	91	2.2	356	1	VF39_NPVLD	MAJOR CAPSID PROTEIN.	4.44e+00
22	93	2.2	380	1	DP3B_MYCPN	DNA POLYMERASE III, BE	2.33e+00
23	92	2.2	442	1	CEM1_YEAST	3-OXOACYL--ACYL-CARRIE	3.22e+00

Db	188	IPVIAQCLNDEGGPSLSHRHWNSTELKVELEC-DID-GRSY-R-Q-IHSRTIKTD-ND	241
QY	227	VSRYAQLCRGDQGESLSYSKWNTEFLKAMLVCSDAATNKNRQLQDVLELLPDPQGWDR	286
Db	242	TILVVFEDSPYSKALCTYSMMTIKOSFSTSKLEGVTKOLPSPAPGICLPAGKVVSHTF	301
QY	287	TRYGVFSNPWNVAVCVISGLDIDKVFRTSSLKGYHSSLNPRPCKCLPDQOPIETEF	346
Db	302	EVIEKYNVLDDIITKPLS	318
QY	347	QVADRHPEVAORVEPMG	363

```

RESULT      2
ID          VA39_VACCV      STANDARD:      PRT:      441 AA.
AC          P24764;
DT          01-MAR-1992 (REL. 21, CREATED)
DT          01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT          01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE          PROTEIN A39
DT          DE
GN          A39R OR (SALL9R AND Salfir).
OS          VACCINIA VIRUS (STRAIN WR).
OC          VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC          ORTHOPOXVIRUS.
RN          [1]
RX          SEQUENCE FROM N.A.
RX          MEDLINE: 91310644.
RA          AMEGADZIE B.Y., AHN B.-Y., MOSS B.;
RT          "Identification, sequence, and expression of the gene encoding a Mr
RT          35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase.";
RL          J. BIOL. CHEM. 266:13712-13718(1991).
RN          [2]
RX          SEQUENCE FROM N.A.
RX          MEDLINE: 91259063.
RA          SMITH G.L., CHAN Y.S., HOWARD S.T.;
RT          "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RT          the right inverted terminal repeat.";
RL          J. GEN. VIROL. 72:1349-1376(1991).
CC          -1- CAUTION: IN REF.2 THIS SEQUENCE IS IN TWO PART (SALL9R AND
CC          Salfir) DUE TO THE PRESENCE OF FRAMESHIFT COMPARED TO THAT OF
CC          REF.1.
-----
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration
CC          between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC          the European Bioinformatics Institute. There are no restrictions on its
CC          use by non-profit institutions as long as its content is in no way
CC          modified and this statement is not removed. Usage by and for commercial
CC          entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC          or send an email to license@lsb-sib.ch).
-----
DR          EMBL; M61187; G335798;
DR          EMBL; D11079; G222726; ALT_SEQ.
DR          EMBL; X57318; G62254;
DR          PIR; S29921; S29921.
SQ          SEQUENCE 441 AA; 50185 MW; 2C823A68 CRC32;

```

Query Match	11.0%	Score 463;	DB 1;	Length 441;
Best Local Similarity	33.9%;	Pred. No. 2.81e-84;		
Matches	87:	Conservative	57:	Mismatches 96:
			Indels	17:
			Gaps	16:

Db	115	LVCGTNNGNPKWK -	IGSDDDPKHRGGRGYAPYQNSKVTIISHNGC-VLSDINISK-EG-I	170
		: : : : : : :		
Qy	109	LACGTNARHPSCNLVNVTVP	LGEMRGYAPFSPDENSLVLPESGDEVYSTIRKQYNGKI	168
		: : : : : : :		
Db	171	KWRREDPCGYDLYTADNVIPKDG-	LRGAFTVKDGTVD-KVYLTFDTIGSKR--I-VK	225
		: : : : : : : :		
Qy	169	PFRIRIGE-S-ELYTSDTYMQNPQIKATIVH	QDQDAYDDKIYFFREDNPDKNPEAPLN	226
		: : : : : : : :		
Db	226	IPYAQWCLNDEGGPSSLSHRHSNFTLKV	VELEC-DTD-GRSY-R-Q-IHSRTIKTD-ND	279
		: : : : : : : :		
Qy	227	VSRAVQLCRGDGGESSLSYKNTFLKMLV	GSDAATNKNFRLQDVELLPDPGOWRD	286
		: : : : : : : :		

Db	280	TILYVFDS	SYSKALCTYS	NMNTIKOS	FSFSTK	LEGYTK	QOLPSP	ASGICL	PAGKVV	PHTF	339
Qy	287	TRVGVFS	NPNWYSA	CVCSILG	IDIKVFT	SSLKGH	SSLNPR	PKCLP	DDQOPI	PTET	346
Db	340	EVIEKYN	VLDDII	KPLS	356						
Qy	347	QVADR	HPHVAQ	RVEPMG	363						

```

RESULT      3
ID  SEX_HUMAN  STANDARD;      PRT;   1871 AA.
AC  P31805;
DT  01-OCT-1996 (REL. 34, CREATED)
DT  01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT  01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE  TRANSMEMBRANE PROTEIN SEX PRECURSOR.
GN  SEX.
OS  HOMO SAPIENS (HUMAN).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RC  SEQUENCE FROM N.A.
RC  TISSUE=FETAL BRAIN, SKELETAL MUSCLE, AND EMBRYO;
RX  MEDLINE; 96149362
RA  MAESTRINI E., TAMAGNONE L., LONGATI P., CREMONA O., GULISANO M.,
RA  BIONE S., TAMANINI F., NEEL B.G., TONIOLO D., COMOGGIO P.M.;
RT  "A family of transmembrane proteins with homology to the
RT  MET-hepatocyte growth factor receptor.";
RL  CC-PROT. NATL. ACAD. SCI. U.S.A. 93:674-678(1996).
CC  -1- FUNCTION: PUTATIVE RECEPTOR INVOLVED IN THE DEVELOPMENT OF NEURAL
CC  AND EPITHELIAL TISSUES.
CC  -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC  -1- TISSUE SPECIFICITY: DOMINANT IN NEURAL TISSUE DURING
CC  DEVELOPMENT (BY SIMILARITY); WIDELY EXPRESSED IN ADULT TISSUE.
CC  -1- SIMILARITY: WITH THE EXTRACELLULAR DOMAIN OF THE MET/RON/HGF
CC  RECEPTORS AND THE CYTOPLASMIC DOMAIN OF SEP, NOV AND OCT.
-----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/
CC  or send an email to license@isb-sib.ch).

```

DR	EMBL:	X87852;	E183847;	-.
DR	NM_	300022;	--	
KW	RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.			
FT	SIGNAL	1	19	POTENTIAL.
FT	CHAIN	20	1871	TRANSMEMBRANE PROTEIN SEX.
FT	DOMAIN	20	1220	POTENTIAL.
FT	TRANSMEM	1221	1241	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	1242	1871	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	59	59	POTENTIAL.
FT	CARBOHYD	548	548	POTENTIAL.
FT	CARBOHYD	637	637	POTENTIAL.
FT	CARBOHYD	738	738	POTENTIAL.
FT	CARBOHYD	746	746	POTENTIAL.
FT	CARBOHYD	1009	1009	POTENTIAL.
FT	CARBOHYD	1036	1036	POTENTIAL.
FT	CARBOHYD	1073	1073	POTENTIAL.
FT	CARBOHYD	1115	1115	POTENTIAL.
FT	CARBOHYD	1162	1162	POTENTIAL.
SO	SEQUENCE	1871 AA;	207661 MW;	9A11046A CRC32;

```

Query Match      3.98; Score 163; DB 1; Length 1871;
Best Local Similarity 29.5%; Pred. No. 9.40e-13;
Matches 26; Conservative 29; Mismatches 29; Indels 4; Gaps 3;

Db 427 VFVIGTSGSLKRVVDGQDAHL-YETVPVVD--GSPILRDLLFSPDRHRYLLSEKQV 483
Ov 394 VLYLTDRGTHIKVPEGEHSAFNIMIOFPRAAAIQTMSLDAERKKLYSSQWEV 453

```


QY 131 LGE-MRGYAPFSPDENSESLVFEQDEVYTIKQYNGK-IPRFRIRGESE-LYTSDTYM 187
Db 252 ES 253
QY 188 QN 189

RESULT 6
ID NRT1_RAT STANDARD; PRT; 275 AA.
AC P17982.
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 (T-CELL NAD(P)(+)-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL MONO(ADP-RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN RT6.1.1).
GN ART2A OR RT6-A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEWIS A.
RX MEDLINE; 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell alloantigen RT6.1";
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
RN [2]
RP MUTAGENESIS OF GLN-207.
RX MEDLINE; 96275529.
RA MAEHAMA T., HOSHINO S.-I., KATADA T.;
RT "Increase in ADP-ribosyltransferase activity of rat T lymphocyte alloantigen RT6.1 by a single amino acid mutation";
RL FEBS LETT. 388:189-191(1996).
CC -!- FUNCTION: HAS NAD+ GLYCOHYDROLASE ACTIVITY AND EXTREMELY LOW ADP-RIBOSYLTRANSFERASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE + N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -!- CATALYTIC ACTIVITY: NAD(+) + H(2)O = NICOTINAMIDE + ADP-RIBOSE.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -!- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; X52082; G57168; -.
CC DR EMBL; M31138; G206804; -.
CC DR PIR; S08464; S08464.
CC DR PROSITE; PS01291; ART; 1.
CC DR PFAM; PF01129; ART; 1.
CC KW TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
CC T-CELL DIFFERENTIATION; GPI-ANCHOR.
FT SIGNAL 1 20
FT CHAIN 21 246 T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
FT PROPEP 247 275 HYDROPHOBIC, REMOVED DURING MATURATION (BY SIMILARITY).
FT
FT LIPID 246 246 GPI-ANCHOR (BY SIMILARITY).
FT ACT_SITE 209 209 BY SIMILARITY.
FT CARBOHYD 58 58 POTENTIAL.
FT MUTAGEN 207 207 Q->E; INCREASED ADP-RIBOSYLTRANSFERASE ACTIVITY.
SQ SEQUENCE 275 AA; 31388 MW; 523B1A84 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 3.12e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;
Db 194 LGVYIKERS-FYPDQEE-VLIPGYEVYQKVRTOGYNEIFLDSFKRKKSNYNCLYSAGTR 251
QY 131 LGE-MRGYAPFSPDENSESLVFEQDEVYTIKQYNGK-IPRFRIRGESE-LYTSDTYM 187
Db 252 ES 253
QY 188 QN 189

RESULT 7
ID EF1A_TRIE STANDARD; PRT; 460 AA.
AC P34825.
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEFL.
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OM9414 / RUTC-30;
RX MEDLINE; 94124021.
RA NAKARI T., ALATALO E., PENTTILA M.;
RT "Isolation of Trichoderma reesei genes highly expressed on glucose-containing media: characterization of the tefl gene encoding translation elongation factor 1 alpha";
RL GENE 136:313-318(1993).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; Z23012; G312887; -.
CC DR PIR; S35772; S35772.
CC DR PROSITE; PS00301; EFATOR_GTP; 1.
CC DR PFAM; PF00009; GTP_EFTU; 1.
CC DR HSP; P07157; IALP.
CC KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 49830 MW; B9ABAB2 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 3.12e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;
Db 210 YKGWKEKFKAGFTG-KILLEAIDSIIEPKR-PTDKPLRLPQDV 252
QY 478 YCGWDQGRICISYSSERSVLQINPAEPKPCPNPKDKAPLQKV 522

RESULT 8
ID VRP2_SALTY STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
65 KD VIRULENCE PROTEIN.
GN MKAA
OS SALMONELLA TYPHIMURIUM.
OC PLASMID 96 KB VIRULENCE PEK102.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-TML R66;
RX MEDLINE: 90136009.
RA TAIRA S., RHEN M.;
RT Identification and genetic analysis of mkaA -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth.";
RL MICROB. PATHOG. 7:165-173(1989).
RN [2]
RP SEQUENCE OF 1-10.
RX MEDLINE: 91244158.
RA TAIRA S., BAUMANN M., RIIKONEN P., SUKUPOLVI S., RHEN M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium.";
RL FEMS MICROBIOL. LETT. 61:319-323(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z15042; G47783; -
DR PIR: A54540; A54540.
KW PLASMID: VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65440 MW; B69F25C3 CRC32;
Query Match 2.4%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.12e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLGKTAARLSDPOAASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RCENYVITLLRSEGLACGTNARHPSCNVLNGTVPLGEMRGYAPFSPDENSILVFE 151
Db 220 GNEAGDRSAMRYSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRIRGESELY 181

RESULT 9
ID VRP2_SALCH STANDARD; PRT; 591 AA.
AC P17450;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN (PROTEIN M2, IN MBA REGION).
OS SALMONELLA CHOLERAE-SUIS (SALMONELLA ENTERICA).
OG PLASMID PKDSC50.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-RF-1;
RX MEDLINE: 90245675.
RA MATSUI H.;
RT "Nucleotide sequences of genes encoding 32 kDa and 70 kDa
RT polypeptides in mba region of the virulence plasmid, pKDC50, of

Salmonella choleraesuis.";
RL NUCLEIC ACIDS RES. 18:2181-2181(1990).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X52035; G46898; -
DR PIR: S09498; S09498.
KW PLASMID: VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65341 MW; BEA4A39A CRC32;
Query Match 2.4%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 3.12e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLGKTAARLSDPOAASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RCENYVITLLRSEGLACGTNARHPSCNVLNGTVPLGEMRGYAPFSPDENSILVFE 151
Db 220 GNEAGDRSAMRYSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRIRGESELY 181

RESULT 10
ID VRP2_SALEN STANDARD; PRT; 591 AA.
AC P55220;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN SPVB.
OS SALMONELLA ENTERITIDIS.
OG PLASMID PNL2001.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-AL1190;
RX MEDLINE: 94362897.
RA SUZUKI S., KOMASE K., MATSUI H., ABE A., KAWAHARA K., TAMURA Y.,
RA KIJIMA M., DANBARA H., NAKAMURA M., SATO S.;
RT "Virulence region of plasmid PNL2001 of Salmonella enteritidis.";
RL MICROBIOLOGY 140:1307-1318(1994).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D14490; G517164; -
DR PIR: S09498; S09498.
KW PLASMID: VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65350 MW; 9C5B2EE3 CRC32;
Query Match 2.4%; Score 99; DB 1; Length 591;

```
Best Local Similarity 23.3%; Pred. No. 3.12e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHAQW-LVEESTPAGE-HIYISYLAENGDNVDLN 219
QY 92 RDCENYITLLERSEGLLACGTNARHPSCWNLVGTVPGLGEMRGYAPFSPDENSELVLE 151
Db 220 GNEAGDRSAMRYSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRIRGESELY 181

RESULT 11
ID VRP2_SALDU STANDARD; PRT; 593 AA.
AC P24419;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN VSDC.
OS SALMONELLA DUBLIN.
OG PLASMD PSDL2.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC SALMONELLA.
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE: 91251759.
CC KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINEY D.;
CC "Molecular analysis of the virulence locus of the Salmonella dublin
CC plasmid PSDL2.";
CC MOL. MICROBIOL. 5:307-316(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X56727; G47839; -.
CC PIR: S15215; S15215.
CC KW PLASMD; VIRULENCE.
CC FT DOMAIN 367 375 POLY-PRO.
CC SQ SEQUENCE 593 AA; 65613 MW; A7CE57B5 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 593;
Best Local Similarity 23.3%; Pred. No. 3.12e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHAQW-LVEESTPAGE-HIYISYLAENGDNVDLN 219
QY 92 RDCENYITLLERSEGLLACGTNARHPSCWNLVGTVPGLGEMRGYAPFSPDENSELVLE 151
Db 220 GNEAGDRSAMRYSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKQYNGKIPFRIRGESELY 181

RESULT 12
ID PMAL_AJECA STANDARD; PRT; 916 AA.
AC Q07421;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).
GN PMAL.
OS AJELLOWYCES CAPSULATA (HISTOPLASMA CAPSULATUM).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
```

```
ONYGENALES; ONYGENACEAE; AJELLOWYCES.
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE: 94124018.
CC SCHAFER M.P., DEAN G.E.;
CC "Cloning and sequence analysis of an H(+)-ATPase-encoding gene from
CC the human dimorphic pathogen Histoplasma capsulatum.";
CC GENE 136:295-300(1993).
CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING
CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKINIZATION MAY MEDATE
CC GROWTH RESPONSES.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (F1-F2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L07305; G409249; -.
CC PROSITE: PS00154; ATPASE_E1_E2; 1.
CC PRAM: P000122; E1-E2-ATPase; 1.
CC KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
CC ATP-BINDING.
CC FT DOMAIN 1 111 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 112 134 1 (POTENTIAL).
CC FT DOMAIN 135 136 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 137 155 2 (POTENTIAL).
CC FT DOMAIN 156 279 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 280 305 3 (POTENTIAL).
CC FT DOMAIN 306 320 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 321 350 4 (POTENTIAL).
CC FT DOMAIN 351 711 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 712 734 5 (POTENTIAL).
CC FT DOMAIN 735 754 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 755 776 6 (POTENTIAL).
CC FT DOMAIN 777 817 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 818 843 7 (POTENTIAL).
CC FT DOMAIN 844 847 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 848 874 8 (POTENTIAL).
CC FT DOMAIN 875 916 CYTOPLASMIC (POTENTIAL).
CC FT MOD_RES 374 374 ASP/GLU-RICH (ACIDIC).
CC FT BINDING 470 470 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 916 AA; 98884 MW; 4A68A44 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 916;
Best Local Similarity 37.2%; Pred. No. 3.12e-01;
Matches 16; Conservative 8; Mismatches 16; Indels 3; Gaps 3;

Db 762 LAVGTWIT-LTTHLVGSENGIYQVNGFTHPVLFLEISLTENW 803
QY 244 LSVSKWNTFLKMLVCSO-AATNKNRNLQDV-FLPDPSQW 284

RESULT 13
ID HA21_HUMAN STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RA  MEDLINE; 84186117.
RA  AUFFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPES D.,
RA  STROMINGER J.L.;
RT  "Isotypic and allotypic variation of human class II
RT  histocompatibility antigen alpha-chain genes.";
RL  NATURE 308:327-333(1984).
RN  [2]
RP  SEQUENCE OF 40-355 FROM N.A.
RA  MEDLINE; 83065171.
RA  AUFFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT  "cDNA clone for the heavy chain of the human B cell alloantigen DCI:
RT  strong sequence homology to the HLA-DR heavy chain.";
RL  PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; J00199; -: NOT_ANNOTATED_CDS.
DR  PIR; A02212; HLRUDC.
DR  PROSITE; PS00290; IG_MHC; 1.
DR  PFAM; PF00047; Ig; 1.
DR  PFAM; PF00993; MHC_II_alpha; 1.
DR  HSP; P01910; IIAK.
KW  MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT  SIGNAL 1 23
FT  CHAIN 24 255
FT  DOMAIN 24 255
FT  DOMAIN 111 204
FT  DOMAIN 205 217
FT  DOMAIN 218 240
FT  TRANSMEM 218 240
FT  DOMAIN 241 255
FT  DISULFID 133 189
FT  CARBOHYD 104 104
FT  CARBOHYD 144 144
SQ  SEQUENCE 255 AA; 5646DD59 CRC32;

Query Match 2.3%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1.21e+00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;

Db 25 DIVADVAQLG-VNLYQSGYSGQYSHFEDGDEEFYVDLERKETVQLPFRFR 78
Qy 122 NLVNGTVVPLGEMRGYAPSPDENSELVLFEGDEV-YSTIRKQYNGKIPFRFR 175

RESULT 14
ID VL2_HPV5B STANDARD; PRT; 518 AA.
AC P26540;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MINOR CAPSID PROTEIN L2.
GN L2.
OS HUMAN PAPILLOMAVIRUS TYPE 5B.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA YABE Y., SAKAI A., HITSUMOTO T., KATO H., OGURA H.;
RT "A subtype of human papillomavirus 5 (HPV-5b) and its subgenomic
RT segment amplified in a carcinoma: nucleotide sequences and genomic
RT organizations.";
RL VIROLOGY 183:793-798(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90252; D1015003; -.
DR PIR; H40480; P2WLB5.
DR PFAM; PF00513; late_protein_L2; 1.
KW COAT PROTEIN; LATE PROTEIN.
SQ SEQUENCE 518 AA; 56836 MW; D0ADC243 CRC32;

Query Match 2.3%; Score 95; DB 1; Length 518;
Best Local Similarity 27.1%; Pred. No. 1.21e+00;
Matches 16; Conservative 17; Mismatches 25; Indels 1; Gaps 1;

Db 219 GDITDIIIELEIPSRYTEIEEPTPPRSSTPLPRNQSVGRGFSLTNRRLVQOVVD 277
Qy 402 GTIHKVPEGEHSAFNFMEIQFRRAAAIQTMSLDAERKKLY-VSSQWEVSQVPLD 459

RESULT 15
ID VL2_HPV05 STANDARD; PRT; 518 AA.
AC P06918;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MINOR CAPSID PROTEIN L2.
GN L2.
OS HUMAN PAPILLOMAVIRUS TYPE 5.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ZACHOW K.R., OSTROW R.S., FARAS A.J.;
RT "Nucleotide sequence and genome organization of human papillomavirus
RT type 5.";
RL VIROLOGY 158:251-254(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M17463; G484221; -.
DR PIR; B26277; P2WLB5.
DR PFAM; PF00513; late_protein_L2; 1.
KW COAT PROTEIN; LATE PROTEIN.
SQ SEQUENCE 518 AA; 56704 MW; B7199004 CRC32;

Query Match 2.3%; Score 95; DB 1; Length 518;
Best Local Similarity 27.1%; Pred. No. 1.21e+00;
Matches 16; Conservative 17; Mismatches 25; Indels 1; Gaps 1;

Db 219 GDITDIIIELEIPSRYTEIEEPTPPRSSTPLPRNQSVGRGFSLTNRRLVQOVVD 277
Qy 402 GTIHKVPEGEHSAFNFMEIQFRRAAAIQTMSLDAERKKLY-VSSQWEVSQVPLD 459

Search completed: Thu Jul 8 19:45:19 1999
Job time : 23 secs.

```

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:45:38 1999; MasPar time 37.43 Seconds
825.329 Million cell updates/sec
Tabular output not generated.

Title: >US-09-041-236-2

Description: (1-566) from US09041236.pep (40 of 45)

Perfect Score: 4211

Sequence: 1 LLLLWAAAQAQGHLSGP.....HKENVEQSCPEGHQSPNCIL 566

Scoring table: PAM 150

Gap	11
-----	----

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 48.690: Variance 74.651: scale 0.652

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			ID	Description	Pred. No.
	Score	Match	Length			
1	4211	100.0	666	4	SEMAPHORIN L.	0.00e+00
2	2398	56.9	393	11	SEMAPHORIN L (FRAGMENT	0.00e+00
3	1850	43.9	653	14	SIMILAR TO GENBANK ACC	0.00e+00
4	737	17.5	748	11	SEMAPHORIN A PRECURSOR	2.83e-155
5	725	17.2	751	11	SEMAPHORIN E PRECURSOR	4.39e-152
6	716	17.0	749	4	SEMAPHORIN V.	1.04e-149
7	716	17.0	750	4	SEMAPHORIN V.	1.04e-149
8	710	16.9	751	4	SEMAPHORIN E.	1.04e-148
9	705	16.7	751	13	COLLAPSSIN 3.	4.01e-147
10	688	16.3	753	4	SEMAPHORIN III FAMILY	8.41e-147
11	682	16.2	754	11	SEMAPHORIN III ISOFORM	2.60e-142
12	681	16.2	772	13	SEMAPHORIN IV ISOFORM	9.95e-141
13	682	16.2	785	11	COLLAPSSIN.	1.83e-140
14	681	16.2	785	4	SEMAPHORIN IV ISOFORM	9.95e-141
15	679	16.1	785	4	SEMAPHORIN IV.	1.83e-140
16	669	15.9	772	11	SEMAPHORIN.	6.15e-140
17	664	15.8	772	11	SEMAPHORIN D PRECURSOR	2.65e-137
18	656	15.6	771	4	SEMAPHORIN III/COLLAPS	5.50e-136
19	618	14.7	775	11	SEMAPHORIN-III.	7.00e-134
20	618	14.7	782	11	SEMAPHORIN H.	6.52e-124
					SEMAPHORIN C (SEM C (6.52e-124

```
Db 333 AVCVYSLGDIKVFRTSSSLKGYHSSLPNPRPGKCLPDQOPIPTETFOVADRHPEVAQRVE 392
|||||
QY 301 AVCVYSLGDIKVFRTSSSLKGYHSSLPNPRPGKCLPDQOPIPTETFOVADRHPEVAQRVE 360

Db 393 PMGPKLTPLFHSKYHYQKVAVHRMQASHGETFHVLYLTTRDGTIHKVVEPGEQEHSAFN 452
|||||
QY 361 PMGPKLTPLFHSKYHYQKVAVHRMQASHGETFHVLYLTTRDGTIHKVVEPGEQEHSAFN 420

Db 453 IMEIQFRAAAATQMSLDAERKLYVSQWEVSQVPLDLCVEYGGCGHGLMSRDPYCG 512
|||||
QY 421 IMEIQFRAAAATQMSLDAERKLYVSQWEVSQVPLDLCVEYGGCGHGLMSRDPYCG 480

Db 513 WDQGRICISYSSRSVLSQINPAEPKPCPNKAPKQVSLAPNSRYILSCPMESRH 572
|||||
QY 481 WDQGRICISYSSRSVLSQINPAEPKPCPNKAPKQVSLAPNSRYILSCPMESRH 540

Db 573 ATYSWRHKNVEQSCPEGHQSPNCIL 598
|||||
QY 541 ATYSWRHKNVEQSCPEGHQSPNCIL 566

RESULT 2
ID O88371 PRELIMINARY: PRT: 393 AA.
AC O88371:
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN L (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
GN SEMAL.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030699; G3523117; -.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;

Query Match 56.9%; Score 2398; DB 11; Length 393;
Best Local Similarity 90.4%; Pred. No. 0.00e+00;
Matches 329; Conservative 16; Mismatches 15; Indels 4; Gaps 2;

Db 33 LLLVFWAAASAGHSRSGPRISAVNKG--QDHVDFSQPEHTVLFHEPGSFVWVGGR 89
|||||
QY 1 LLLLLWAAASAGHLRSGPRIFAVKWHGVQDRVDFGOTEPHTVLFHEPGSSVWVGGR 60

Db 90 GKVIYHFEPEGKNASVTRTNIWGTGSCQDQKQCGNYITLLERRGNLLVCGTNARKPSC 149
|||||
QY 61 GKVIYDFPEEGKNASVTRTNIWGTGSCQDQKQCGNYITLLERRSGLLACGTNARHPSC 120

Db 150 WNLVNDVSVMLGEMKGYAPFSDENSLVLFEGDEVYSTIRKOEYNGKIPRFRIRGESE 209
|||||
QY 121 WNLVNGTVV-PLGEMRGYAPFSDENSLVLFEGDEVYSTIRKOEYNGKIPRFRIRGESE 179

Db 210 LYTSVTVMQNPQIKATIVHQDAYDDKIYFFREDNPKNPAPLNVSRAQLCRGDGQ 269
|||||
QY 180 LYTSVTVMQNPQIKATIVHQDAYDDKIYFFREDNPKNPAPLNVSRAQLCRGDGQ 239

Db 270 GESSLSVSKWNTFLKAMLVCSDAATNRNRLQDVLLPDPSCQWRDTRVYGVFSNPNY 329
|||||
QY 240 GESSLSVSKWNTFLKAMLVCSDAATNRNRLQDVLLPDPSCQWRDTRVYGVFSNPNY 299

Db 330 SAVCVYSLGDIKVFRTSSSLKGYHSSLPNPRPGKCLPDQOPIPTETFOVADRHPEVAQRV 389
|||||
QY 300 SAVCVYSLGDIKVFRTSSSLKGYHSSLPNPRPGKCLPDQOPIPTETFOVADRHPEVAQRV 359

Db 390 EPMG 393
```

```
QY 360 EPMG 363
|||||

RESULT 3
ID Q64906 PRELIMINARY: PRT: 653 AA.
AC Q64906;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALCELAAPHINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RX MEDLINE; 97201573.
RA ENSSER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RX MEDLINE; 97404659.
RA ENSSER A., PELANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RA ENSSER A., PELANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997), TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U18243; G1000717; -.
SQ SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match 43.9%; Score 1850; DB 14; Length 653;
Best Local Similarity 46.8%; Pred. No. 0.00e+00;
Matches 267; Conservative 102; Mismatches 189; Indels 13; Gaps 13;

Db 57 LLMIL-SAITAASKSRFIDKPRILVNLTDGFGQHRF-FGPOEPHTVLFHSLNSDDVYVGN 114
|||||
QY 1 LLLLLWAAASAGHLRSGPRIFAVKWHGVQDRVDFGOTEPHTVLFHEPGSSVWVGGR 60

Db 115 NTIYLFDFAHSSNASTALINITSTNTHRLSTCENFTLLHNQTDGLACGTNSOKPSC 174
|||||
QY 61 GKVIYDFPEEGKNASVTRTNIWGTGSCQDQKQCGNYITLLERRSGLLACGTNARHPSC 120

Db 175 W-LINNLITQELGPKGLAPFSPSSGNLVLPDNDTYSTINLYKSLSGS-HKFRRTAGOV 232
|||||
QY 121 WNLVNGTVV-PLGEMRGYAPFSDENSLVLFEGDEVYSTIRK-OEYNGKIPRFRIRGES 178

Db 233 ELYTSDTAHMRPQFVOATAVHKHNSYDDKIYFFFOENSHSDFKQPHPTVPRVQVCSSDQ 292
|||||
QY 179 ELYTSDTYMNPQFIKATIVHQDAYDDKIYFFREDNPKNPAPLNVSRAQLCRGDQ 238

Db 293 GESSLSVSKWNTFLKARLACVYDGTGRYINELQDIFWQAPENSWEETLYLGLFLSPWN 352
|||||
QY 239 GESSLSVSKWNTFLKARLACVYDGTGRYINELQDIFWQAPENSWEETLYLGLFLSPWN 298

Db 353 FSACVCTVKDIDHVFKTSKLNKHHKLPTRPGOCMKNHQVPTETFOVADRYPEVADP 412
|||||
QY 299 YSACVYSLGDIKVFRTSSSLKGYHSSLPNPRPGKCLPDQOPIPTETFOVADRHPEVAQR 358

Db 413 YQKNNAFPPIQSIYITKLLVYVEYG-GVFWATIFYLTTIKGTIHYIYRVEDSNSTT 471
|||||
QY 359 VEPMPGLKTPLFHSKYHYQKVAVHRMQASHGETFHVLYLTTRDGTIHKVVEPGEQEHSE 417

Db 472 ALNILEINPQFQPAIQIONLLDNTNLKLYVNSEWESEVPLDLCVYGNDCSCFMSRDP 531
|||||
QY 418 AFNIMEIOPFRRAAAATQMSLDAERKLYVSQWEVSQVPLDLCVEYGGCGHGLMSRDP 477
```



```
RESULT 8
ID Q99985 PRELIMINARY; PRT; 751 AA.
AC Q99985;
DT 01-MAY-1997 (TREMBREL. 03, CREATED)
DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E. (HUMAN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA YAMADA T., ENDO R., GOTOH M., HIROHASHI S.;
RT "Identification of semaphorin E as a non-MDR drug resistance gene of
human cancers.";
RL PROC. NATL. ACAD. SCI. U.S.A. 94:14716-14718(1997).
DR EMBL; AB000220; D1033360; -.
SQ SEQUENCE 751 AA; 85207 MW; F89FB934 CRC32;

Query Match 16.9%; Score 710; DB 4; Length 751;
Best Local Similarity 33.6%; Pred. No. 4.01e-148;
Matches 151; Conservative 104; Mismatches 151; Indels 43; Gaps 30;

Db 167 SFPNNTVSVMIINEELFSGMYI-DFMGTDAAIFRSILTKRNVARTDOHNSKWLSEPFVD 225
QY 139 PFSFDENSLVLFEGDEVYSTIRKOEYNGK-IPFRIRIGESELYTS--DTV-MONPOFIK 194
Db 226 AHVIPDGTDPNDARVYFFFKELTNNRSTKQIHSIARIICPNNDTGGRLSL-VNKKWTFEL 284
QY 195 ATIVHQD-QAYDDKIYFFREDPNKPEAPLNVSRVAQLCRDGGESSLSVSKNTFL 253
Db 285 KARLVCSVDEDEGPEHFDELEDFLL-ETDNP-RTTLVYGIFTSSSVFKGSVAVCYHL 342
QY 254 KAMLVCS--DA-ATNKNFNRLQDFVLLPDPGOWRDRTRYGVESPNW-Y--SAVCVYSL 307
Db 343 SDIQTVFNPGFAHKEGNHOLISYQGRIPYPRGTCPPGGAFTPMRTTKEFFDDVVFIR 402
QY 308 GDIDKVFRTS-SLK-G-----YHSSLNPRPGKCLPDQ-QP-IPDET-F--QVAD--R 351
Db 403 NHPLMYNPIYPIHKRPLVRI-GTDYKTKIADRVNAADG-RYHVLFLGTDRGTQVKV 460
QY 352 -HPEVAQVEPM-G-PLKTLFHSKYHVKVAVHRQAASHGETHFVLYLTDRGTGTHKV 408
Db 461 VLPTNSVSGELILEEVFNKHAIPITMKISSKKQOLYSSNEGVSQVSLHRCHYIGTA 520
QY 409 E-PGEQHSFAFNIMEIQPFRRAAAIQTMSLDAERRKLYSSQWESQVPLDLCEVYGG 467
Db 521 CADCLARDPYCANDGHSRFFPTGKRSSRQDVRHGNPLTQCRGFNLKAYRNAAEIVQ 580
QY 468 CHGLMSRDPYCGWDOGRCSISYSE--RSVLOSINPAEPHKECP--NPKPDKAPLOKVS 523
Db 581 YGKNNNTFLECAPKSPQASIKWLLQDK 609
QY 524 LAP-NSRYILSCPMESRHATYSNR-HKEN 550

RESULT 9
ID Q42236 PRELIMINARY; PRT; 751 AA.
AC Q42236;
DT 01-JAN-1998 (TREMBREL. 05, CREATED)
DT 01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE COLLAPLIN 3.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA FEINER L., KOPPEL A.M., KOBAYASHI H., RAPER J.A.;
RL NEURON 19:0-0(1997).
DR EMBL; AF022946; G2522204; -.
SQ SEQUENCE 751 AA; 85432 MW; 0B9BA2F7 CRC32;
```

```
Query Match 16.7%; Score 705; DB 13; Length 751;
Best Local Similarity 33.5%; Pred. No. 8.41e-147;
Matches 150; Conservative 107; Mismatches 150; Indels 41; Gaps 28;

Db 167 SFPNNTVSVMIINEELFSGMYI-DFMGTDAAIFRSILTKRNVARTDOHNSKWLSEPFVD 225
QY 139 PFSFDENSLVLFEGDEVYSTIRKOEYNGK-IPFRIRIGESELYTS--DTV-MONPOFIK 194
Db 226 AHVIPDGTDPNDARVYFFFKELTNNRSTKQIHSIARIICPNNDTGGQRLS-VNKKWTFEL 284
QY 195 ATIVHQD-QAYDDKIYFFREDPNKPEAPLNVSRVAQLCRDGGESSLSVSKNTFL 253
Db 285 KARLVCSVDEDEGTEYFDELEDFLL-ETDNP-RTTLVYGIFTSSSVFKGSVAVCYHL 342
QY 254 KAMLVCS--DA-ATNKNFNRLQDFVLLPDPGOWRDRTRYGVESPNW-Y--SAVCVYSL 307
Db 343 SDIQTVFNPGFAHKEGNHOLIPYQGRIPYPRGTCPPGGAFTPMRTTKEFFDDVVFIR 402
QY 308 GDIDKVFRTS-SLK-G-----YHSSLNPRPGKCLPDQ-QP-IPDET-F--QVAD--RHP 353
Db 403 NHPLMYNPIYPIHKRPLVRIIGTDYKTKIADRVNAADGRC-HVLFLGTDRGTQVKVV 461
QY 354 EVAQVEPMGPL-KTPLF-H--SKYHYQVAVHRQAASHGETHFVLYLTDRGTGTHKVE 409
Db 462 LPTNFSASGELILEEVFOSNPIITMKISSKKQOLYSSSEGVTVQVPLHRCRIYGTAC 521
QY 410 -PGEQHSFAFNIMEIQPFRRAAAIQTMSLDAERRKLYSSQWESQVPLDLCEVYGGG 468
Db 522 ADCCLARDPYCANDGHSRFFPTGKRSSRQDVRHGNPLTQCRGFNLKAYRNAAEIVQY 581
QY 469 HGCLMSRDPYCGWDOGRCSISYSE--RSVLOSINPAEPHKECP--NPKPDKAPLOKVS 524
Db 582 GVKNNNTFLECPKSPQASIKWLLQDN 609
QY 525 AP-NSRYILSCPMESRHATYSNR-HKEN 550

RESULT 10
ID Q13372 PRELIMINARY; PRT; 753 AA.
AC Q13372;
DT 01-NOV-1996 (TREMBREL. 01, CREATED)
DT 01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN III FAMILY HOMOLOG.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 96230324.
RA XIANG R.H., HENSEL C.H., GARCIA D.K., CARLSON H.C., KOK K., DALY M.C.,
RA KERACHER K., DEN BERG A., VELDHUIS P., BUYS C.H., NAYLOR S.L.;
RT "Isolation of the human semaphorin III/F gene (SEMA3F) at chromosome
3p21, a region deleted in lung cancer.";
RL GENOMICS 32:39-48(1996).
DR EMBL; U38276; G1061351; -.
DR PFAM; PF00047; ig; 1.
SQ SEQUENCE 753 AA; 84941 MW; BECBEBB0 CRC32;

Query Match 16.3%; Score 688; DB 4; Length 753;
Best Local Similarity 31.4%; Pred. No. 2.60e-142;
Matches 139; Conservative 121; Mismatches 140; Indels 42; Gaps 30;

Db 171 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRLGKQTAMRTQYNSRWLNDSEFIH 229
QY 139 PFSFDENSLVLFEGDEVYSTIRKOEYNGK-IPFRIRIGESELYTS--DTV-MONPOFIK 194
Db 230 AELIPDSEANDKLYFFFRERSAE-APOSPAVYARIGRICLNDGGCHCL-VNKKWTFELK 287
QY 195 ATIVHQDQAYDDKIYFFREDPNKPEAPLNVSRVAQLCRDGGESSLSVSKNTFLK 254
```



```
DE SEMAPHORIN IV ISOFORM B.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ECKHARDT F., MEYERHANS A.;
RT "Molecular cloning and expression pattern of a murine semaphorin
RT homologous to H-sema IV."
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF080090; G337766; -.
SQ SEQUENCE 785 AA; 88493 MW; 6FFCD8F5 CRC32;

Query Match 16.2%; Score 682; DB 11; Length 785;
Best Local Similarity 31.8%; Pred. No. 9,95e-141;
Matches 141; Conservative 119; Mismatches 140; Indels 43; Gaps 31;

Db 202 PYDPKLDATASALINEELYAGVYI-DFMGTDAAIFRLTGKQTAMRTDOYNSRWLNDPSFTH 260
QY 139 PESPENSLVLFEGDEVYSTIRKQYNGK-IPFRIRGESELYTS--DTV-MQNPFQFIK 194
Db 261 AELIPDSARNDDKLYFFFRERSAE-APQNPAYVARIGRICLNDGGHCL-VNKMSTFL 318
QY 195 ATIVHQD-QAYDDKIYFFREDNPKNPEAPLNSRVAQLCRGDOGGESSLSVSKWNTFL 253
Db 319 KARLVCSVPGEDGIEHFELODFV-QQTQDI-RNPVIYAVFTSSGVSFRGSVCVYSM 376
QY 254 KAMLVCS---DAATNKNFNLQDVFLLPDPGQWRDTRVYGFNSP-WNY--SVCVYSL 307
Db 377 ADIRVFNPGFAHKGPYQWMPFSGKMPYPRPGTCPGGTFTPSMKSTKDYDEVINEMR 436
QY 308 GUIDKVERTS-SLK-G--YH-----SS-LPNPRPGKLPDQ-OP-I-PTETF--QVAD--R 351
Db 437 THPLMYQAVYPIQRRLVVRT-GAPYRLTTIAVDQVDAADG-RYEVLFLGTDRGTQVKYI 494
QY 352 -HPEVAQRVEPMG--PLKTPLFHSKYHYQKVAHRMQASHGETFHVLYLTDRGTIHKVV 408
Db 495 VLPKDDOEELMLEBEVEVFKPAPVKTMTSSKRQOLYASAVGTHLSLRCAQYAGAA 554
QY 409 E-PGEOHSAF-NIMEIQPFRRAAAIQTMSLDAERRKLYVSQWEVSVQVPLDLCEYVGG 467
Db 555 CADCCCLARDPYCAWDQACSRVYTSKRRSRQDVRHGNPIRQCGFNSNANKNAVESVQ 614
QY 468 CHGCLMSRDYPYCNQDGRGIS-IYSSER-SVLQSNPAEPHKECP--NPKPKAPLQKVS 523
Db 615 YGVAGSAAFLECPQRSPOATVKW 637
QY 524 LA-PNSRYLSCPMESRHATYSW 545

RESULT 14
ID Q13275 PRELIMINARY; PRT; 785 AA.
AC Q13275;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN IV.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT cancer deletion region and demonstrate distinct expression
RT patterns.";

DE SEMAPHORIN IV.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA NELSON J., BIEWALD T.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 394-436 FROM N.A.
RC TISSUE-PLACENTA;
RX MEDLINE; 96210603.
RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT cancer deletion region and demonstrate distinct expression
RT patterns.";

RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR EMBL; AC000063; G1689373; -.
DR EMBL; U32172; G995788; -.
DR EMBL; U32171; G995786; -.
DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 785 AA; 88381 MW; CD175765 CRC32;

Query Match 16.2%; Score 681; DB 4; Length 785;
Best Local Similarity 31.4%; Pred. No. 1.83e-140;
Matches 139; Conservative 122; Mismatches 139; Indels 43; Gaps 31;

Db 202 PYDPKLDATASALINEELYAGVYI-DFMGTDAAIFRLTGKQTAMRTDOYNSRWLNDPSFTH 260
QY 139 PESPENSLVLFEGDEVYSTIRKQYNGK-IPFRIRGESELYTS--DTV-MQNPFQFIK 194
Db 261 AELIPDSARNDDKLYFFFRERSAE-APQNPAYVARIGRICLNDGGHCL-VNKMSTFL 318
QY 195 ATIVHQD-QAYDDKIYFFREDNPKNPEAPLNSRVAQLCRGDOGGESSLSVSKWNTFL 253
Db 319 KARLVCSVPGEDGIEHFELODFV-QQTQDV-RNPVIYAVFTSSGVSFRGSVCVYSM 376
QY 254 KAMLVCS---DAATNKNFNLQDVFLLPDPGQWRDTRVYGFNSP-WNY--SVCVYSL 307
Db 377 ADIRVFNPGFAHKGPYQWMPFSGKMPYPRPGTCPGGTFTPSMKSTKDYDEVINEMR 436
QY 308 GUIDKVERTS-SLK-G--YH-----SS-LPNPRPGKLPDQ-OP-I-PTETF--QVAD--R 351
Db 437 SHPLMYQAVYPIQRRLVVRT-GAPYRLTTIAVDQVDAADG-RYEVLFLGTDRGTQVKYI 494
QY 352 -HPEVAQRVEPMG--PLKTPLFHSKYHYQKVAHRMQASHGETFHVLYLTDRGTIHKVV 408
Db 495 VLPKDDOEELMLEBEVEVFKPAPVKTMTSSKRQOLYASAVGTHLSLRCAQYAGAA 554
QY 409 E-PGEOHSAF-NIMEIQPFRRAAAIQTMSLDAERRKLYVSQWEVSVQVPLDLCEYVGG 467
Db 555 CADCCCLARDPYCAWDQACSRVYTSKRRSRQDVRHGNPIRQCGFNSNANKNAVESVQ 614
QY 468 CHGCLMSRDYPYCNQDGRGIS-IYSSER-SVLQSNPAEPHKECP--NPKPKAPLQKVS 523
Db 615 YGVAGSAAFLECPQRSPOATVKW 637
QY 524 LA-PNSRYLSCPMESRHATYSW 545

RESULT 15
ID Q15704 PRELIMINARY; PRT; 785 AA.
AC Q15704;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-BRAIN;
RX MEDLINE; 96226360.
RA ROCHE J., BOLDOG F., ROBINSON M., ROBINSON L., VARELLA-GARCIA L.,
RA SWANTON M., WAGGONER B., FISHEL R., FRANKLIN W., GEMMILL R.,
RA DRABKIN H.;
RT "Distinct 3p21.3 deletions in lung cancer and identification of a new
RT human semaphorin.";
RL ONCOGENE 12:1289-1297(1996).
DR EMBL; U33920; G100207; -.
DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 785 AA; 88385 MW; 9971F14E CRC32;

Query Match 16.1%; Score 679; DB 4; Length 785;
Best Local Similarity 31.6%; Pred. No. 6.15e-140;
Matches 140; Conservative 121; Mismatches 139; Indels 43; Gaps 31;

Db 202 PYDPKLDATASALINEELYAGVYI-DFMGTDAAIFRLTGKQTAMRTDOYNSRWLNDPSFTH 260
```

Qy 139 PFSDENSLVLFEGDEVYSTIRKOEYNGK-IPRFRIRGESELYTS--DTV-MQNPOFIK 194
Db 261 AELIPDSAEERNDKLYFFFRERSAE-APQSPAVYARIGRICLNDDGGHCCL-VNKNWSTFL 318
Qy 195 ATIVHOD-QAYDDDKIYFFREDNPKNPEAPLNYSRVAQLCRGOGGESSLSVSKWNTFL 253
Db 319 KARLVCSVPGEDGIETHFDELQDVV-QOTODV-RNPVIYAVFTSSGVSFRGSVAVCVYSM 376
Qy 254 KAMLVCS---DAATNKNFNRLQDVFLLPDPSCQWRDTRVYGVFSNP-WNY--SACVYISL 307
Db 377 ADIRVFNPGPFAHKEGPNYQMPFSGKMPYPRPGTCPCGTFTPSMKSTKDYDPDEVINMR 436
Qy 308 GDIDKVFRTS-SLK-G-YH----SS-LPNRPRGKCLPDQ-OP-I-PTETF--QVAD--R 351
Db 437 SHPLMYQAVYPLQRRPLVVRT-GAPYRLTTIADVQVDAGDG-RYEVLFGLGTDRGTVOKVI 494
Qy 352 -HPEVAORVEPMG--PLKTPLFHSKYHYQYAVHRMQASHGETEHVLYLTDRGTIHKV 408
Db 495 VLPKDDQEMEELMEEVEFKDPAPVKMTMTISSKRQOLYVASAVGVTHLSLHRCQAYGAA 554
Qy 409 E-PGEQEHSAFNIMEIQPFRAAAIQTMSLDAERKLYVSSQWEVSQVPLDLCEVYGGG 467
Db 555 CADCCLARDPYCAWDGQACSRYTASSKRRRRQDVHRGNPIRQCRGFNSNANKNAVESYQ 614
Qy 468 CHGCLMSRDPYCGWDQGRCS-ITYSSER-SVLQSIINPAEPHKECP--NPKPDKAPLQKVS 523
Db 615 YGVAGSAAFLCQPRSPQATVKW 637
Qy 524 LA-PNSRYLSCPMESRHRATYSW 545

Search completed: Thu Jul 8 19:46:22 1999
Job time : 44 secs.

W P E R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:51:05 1999; MasPar time 28.88 Seconds
415.383 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (24-587) from US09041236.ppp (41 of 45)
Perfect Score: 4201
Sequence: 1 AVWKGVGDVDFGQTEPH.....IENLTAQYGHYFCEAQEGS 564
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.257; Variance 152.226; scale 0.238

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	656	15.6	771	13	Human semaphorin III	7.56e-51
2	615	14.6	477	13	Human collapsin.	8.08e-47
3	608	14.5	775	33	Human semaphorin.	3.92e-46
4	606	14.4	861	22	Mouse CD100 antigen.	6.16e-46
5	599	14.3	861	32	Human semaphorin.	2.99e-45
6	544	12.9	862	22	Human CD100 antigen.	7.04e-40
7	512	12.2	776	32	Rat semaphorin W.	9.09e-37
8	463	11.0	441	13	Vaccinia virus semaph	4.97e-32
9	459	10.9	730	13	Grasshopper semaphori	1.21e-31
10	428	10.2	587	32	Human semaphorin W.	1.14e-28
11	390	9.3	974	33	Human secreted protei	6.65e-24
12	378	9.0	712	13	Tribolium semaphorin	7.32e-23
13	367	8.7	650	13	Drosophila semaphorin	2.17e-22
14	362	8.6	724	13	Drosophila semaphorin	1.90e-21
15	352	8.4	930	32	Human semaphorin Y.	6.47e-19
16	325	7.7	888	25	Human semaphorin Z.	

17	316	7.5	887	25	W19856	Rat semaphorin Z.	4.47e-18
18	296	7.0	929	32	W57259	Rat semaphorin Y.	3.20e-16
19	157	3.7	122	13	R71385	Variola major virus s	6.72e-04
20	99	2.4	591	4	R23006	Protein transcribed f	2.29e-01
21	101	2.4	2329	25	W25038	Partial BRCA2 cancer	1.65e-01
22	98	2.3	832	33	W61092	Taq DNA polymerase I	2.71e-01
23	96	2.3	832	33	W61090	Taq DNA polymerase I	3.75e-01
24	91	2.2	478	1	R04881	Recombinant elastase	8.41e-01
25	91	2.2	598	9	R48631	Sequence of nuclear r	8.41e-01
26	92	2.2	832	33	W61091	Taq DNA polymerase I	7.17e-01
27	92	2.2	854	30	W56309	Class II S-receptor ki	7.17e-01
28	92	2.2	855	6	R29815	S receptor kinase pro	7.17e-01
29	93	2.2	4572	30	W52845	A. mediterranei rifam	6.10e-01
30	89	2.1	143	25	W25052	BRCA2 cancer suscepti	1.15e-02
31	89	2.1	534	25	W25031	Partial BRCA2 cancer	1.15e-02
32	90	2.1	638	5	R03924	E. coli HSP (dnak).	9.85e-01
33	89	2.1	680	4	R23143	Mutant thermostable D	1.15e-02
34	89	2.1	757	4	R23141	Mutant thermostable D	1.15e-02
35	89	2.1	787	4	R23142	Mutant thermostable D	1.15e-02
36	89	2.1	832	13	R76691	DNA-polymerase REM-T3	1.15e-02
37	89	2.1	832	12	R64272	T. aquaticus DNA-poly	1.15e-02
38	89	2.1	832	1	P90556	Purified native therm	1.15e-02
39	89	2.1	832	33	W61089	Taq DNA polymerase I	1.15e-02
40	89	2.1	832	13	R76690	Taq DNA-polymerase RE	1.15e-02
41	89	2.1	833	27	W24211	Cleavase DN nuclease.	1.15e-02
42	89	2.1	833	27	W24212	Cleavase DA nuclease.	1.15e-02
43	89	2.1	833	27	W24213	Cleavase DV nuclease.	1.15e-02
44	89	2.1	3418	26	W19211	Human breast cancer s	1.15e-02
45	89	2.1	3418	26	W23287	Human breast and ovar	1.15e-02

ALIGNMENTS

RESULT 1
ID R71380 standard; Protein; 771 AA.
AC R71380; 1995 (first entry)
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB; Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 60-63; 101pp; English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II,
CC (R7444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor binding
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA;

Query Match 15.6%; Score 656; DB 13; Length 771;
Best Local Similarity 31.2%; Pred.No. 7.56e-51;
Matches 143; Conservative 115; Mismatches 155; Indels 45; Gaps 35;
Db 166 rgkspqdkllatllidgelysgta-adfngdrfaifrtghhhpirtedhsrwnldp 224

```
QY 135 RGAPSPDENSEVLFGEDGVYTIKQYNGK-IPFRRIIRGESELYTS--DTV-MQNP 190
Db 225 kfisahisesdpedkvyyffrenaidgehsqkatharigickndfgghrsl-vnkw 283
QY 191 QFTKATIVHO--DQAYDDKIYFFREDNPDKNPEAPLVNSRVAQLCRGDQGGESSLSVSKW 249
Db 284 ttfkarlicsvpgpigidthfdelgdvflm-nfkdp-knpvvygvyftssnifkgsavc 341
QY 250 NTEFLKAMLVCS--DA-ATNKNENRLQDVFLLPDPGQWRDTRVYGVFSNPNW-Y--SAVC 303
Db 342 mymsdvrrvflgpyahrdpnyqgyvpyqrvpypgtcpsktfgfdestklpddvit 401
QY 304 VYSLGIDIKVFTS-SLK-G--YH--S-S-LPNPRGKCLPDQ-QPI-PTETF--QV-A 349
Db 402 farshpamypvfmnnrpiviktvdnyqftgiqvdrdaedgq-ydvvmfigtdvgtvik 460
QY 350 -DR-HPEVAQRVPMGPKLTPL-FHSKYHYQKVAVHRMQASHGETFHVLTTRDGTIHK 406
Db 461 vvaipketwydleevleentvfreptaisamelstkgqlylgstagvaqlphrcdii 520
QY 407 VVE-PCGEQHSFAFIME-IQPFERRAAAIQMSLDAERRKLYVSSQWEVSQVPLDLCEVY 464
Db 521 gkacaeclardycawdgacsacryfptakrrtrrdingdpithcsdlhhdnhghshp 580
QY 465 GGCCHCLMSRDPYCGWDQGRCSIIYSSE-RSVL-QSINPAEPHKECPNPKPKD-----AP 518
Db 581 eerilygvnsstflescsqralvy-wqfqrneer 617
QY 519 LQKVSLA-PMSRYLSC-PMESRHATYSWR-HKENVEQ 553
```

RESULT 2

```
ID R74175 standard; Protein; 477 AA.
AC R74175;
DT 01-NOV-1995 (first entry)
DE Human semaphorin.
KW Collapsin; antibody; therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT binding_site 9...19 "antibody binding site"
FT binding_site 51..65 "antibody binding site"
FT US5416197-A.
PD 16-MAY-1995.
PR 15-OCT-1993; 136922.
PA (UYPE-) UNIV PENNSYLVANIA.
PI Luo Y, Rapar JA;
DR N-PSDB: Q92331.
PT New antibody to human collapsin - used to inhibit the activity of
PT collapsin, to induce neurite out-growth and to treat individuals with
PT nerve damage.
PS Claim 2; Columns 15-18; 11pp; English.
CC An antibody capable of specifically binding at least a portion of
CC the collapsin protein can be used to purify human collapsin and
CC to inhibit the activity of the protein. It can be used to induce
CC neurite outgrowth by neuronal cells and to treat individuals
CC suffering from nerve damage.
SQ Sequence 477 AA;
```

Query Match 14.6%; Score 615; DB 13; Length 477;
Best Local Similarity 32.3%; Pred. No. 8.08e-47;
Matches 126; Conservative 99; Mismatches 126; Indels 39; Gaps 30;

```
Db 1 lehndpeddkvyffrenaidgehsqkatharigickndfgghrsl-vnkwttfklkarl 59
QY 198 VHODQAYDDKIYFFREDNPDKNPEAPLVNSRVAQLCRGDQGGESSLSVSKWTEFLKAML 257
Db 60 lcsvpgpngldthfdelgdvflm-nfkdp-knpvvygvyftssnifkgsavcmysmsdvr 117
```

```
QY 258 VCS--DA-ATNKNENRLQDVFLLPDPGQWRDTRVYGVFSNPNW-Y--SAVCVYSLGDD 311
Db 118 ryflgpyahrdpnyqgyvpyqrvpypgtcpsktfgfdestklpddvitfarshpam 177
QY 312 KVFRTS-SLK-G--YH--S-S-LPNPRGKCLPDQ-QPI-PTETF--QV-A-DR-HPEV 355
Db 178 ynpvfmnnrpiviktvdnyqftgiqvdrdaedgq-ydvvmfigtdvgtvlkvvsipket 236
QY 356 AQRVPMGPKLTPL-FHSKYHYQKVAVHRMQASHGETFHVLTTRDGTIHKVVE-PGEQ 413
Db 237 wydleevlleentvfreptaisamelstkgqlylgstagvaqlphrcdiygkacaecc 296
QY 414 EHSFAFIME-IQPFERRAAAIQMSLDAERRKLYVSSQWEVSQVPLDLCEVYGGCHGCL 472
Db 297 lardpycawdgacsacryfptakrrtrrdingdpithcsdlhhdnhghshspeerilygv 356
QY 473 MRRDPYCGWDQGRCSIIYSSE-RSVL-QSINPAEPHKECPNPKPKD-----APLQKVS 525
Db 357 ensstflescsqralvy-wqfqrneer 385
QY 526 PMSRYLSC-PMESRHATYSWR-HKENVEQ 553
```

RESULT 3

```
ID W63748 standard; Protein; 775 AA.
AC W63748;
DT 01-OCT-1998 (first entry)
DE Human semaphorin.
KW Human; semaphorin; diagnosis; nerve disease; nerve growth inhibitor;
KW neurological disease; atopic skin inflammation; autoimmune disease;
KW pain.
OS Homo sapiens.
PN WO9822504-A1.
PD 28-MAY-1998.
PR 12-NOV-1997; J04111.
PR 15-NOV-1996; JP-321068.
PA (SUNU ) SUMITOMO PHARM CO LTD.
PI Furuyama T, Inagaki S;
DR WPI: 98-312416/27.
DR N-PSDB: V35367.
PT Gene encoding new semaphorin nerve growth inhibitor - useful in
PT diagnosis, treatment and study of neurological diseases
PS Claim 1; Page 33-37; 49pp; Japanese.
CC The present sequence represents human semaphorin, a nerve growth
CC inhibitor. The semaphorin protein, and gene encoding the protein,
CC and their derivatives, are used in the diagnosis, treatment and
CC study of neurological disorders such as atopic skin inflammation,
CC autoimmune diseases and pain.
SQ Sequence 775 AA;
```

Query Match 14.5%; Score 608; DB 33; Length 775;
Best Local Similarity 32.1%; Pred. No. 3.92e-46;
Matches 125; Conservative 98; Mismatches 130; Indels 36; Gaps 28;

```
Db 167 rgcrpdpnssfvstlvgnellfaglysv-dywdgrdaifrmgkglhrttehdderllkep 225
QY 135 RGAPSPDENSEVLFGEDGVYTIKQYNGK-IPFRRI-R-GE-SELYTSDTVMQNP 190
Db 226 kfvgymipdnedrddnkmyffftkaleaennahtiytrvgrlcvmndggqril-vnkw 284
QY 191 QTIKATIVHQDQAYDD-KIYFFREDNPDKNPEAPLVNSRVAQLCRGDQGGESSLSVSKW 249
Db 285 stflkarlicsvpgpmgidtyfdeledvflp-trdp-knpvifglfnttsnifrhavc 342
QY 250 NTEFLKAMLVCSDAATN-KN--FNRLQDVFLLPDPGQWRDTRVYGVFSNPNW-Y-S-AVC 303
Db 343 vyhmsiarefngpyahkegpyhswlyegkvypprpgscaskvngkgvgttkdydpdai 402
QY 304 VYSLGIDIKVFTS-SLK-G--YHSSL-----PNRPGKCLPD-QQPI--PTETF-QVAD 350
Db 403 rfarmhplmyqkpkvhhkpllvktgdgkynrlqlavdrveaedgq-ydvlfigtgtgvl 461
QY 351 RHPEVAQRV-EPMGPL-KTP-LFHS--KYHYQKVAVHRMQASHGETFHVLTTRDGTIHK 405
```


Db 462 kvitiynqetwemeveileelqifkdpapismissskqlyigsasavaqvrffhcdm 521
 QY 406 KVVPEQGEHSAFN-IME-IQFRRRAAIQIOTMSLDAERKLYVSSQWEVSQVPLDLCEV 463
 Db 522 ygsacacclardpccawdgiscryypt 550
 QY 464 YGGGCHGLMSRDPYCGWDGGRGCRISYSS 492

RESULT

ID W17658 standard; Protein; 861 AA.
 AC W17658;
 DT 24-JUL-1997 (first entry)
 DE Mouse CD100 antigen.
 KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
 KW vaccine.
 OS Mus sp.
 FH Key
 FT peptide
 FT protein
 FT domain
 FT domain
 FT domain
 FT domain
 FT domain
 FT domain
 FT domain
 FT modified_site
 FT W09717368-A1.
 PN 15-MAY-1997. U18645.
 PD 12-NOV-1996; US-556422.
 PR 09-NOV-1995; US-556422.
 PA (DAND) DANA FARBEN CANCER INST.
 PI Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JI;
 DR WPI: 97-280982/25.
 DR N-PSDB; T60665.
 PT Nucleic acid molecule encoding CD100 antigen - which stimulates
 PT leukocyte response, e.g. B cell aggregation, differentiation,
 PT survival and T cell proliferation
 PS Example 8; Page 86-89; 135pp; English.
 CC Mouse CD100 antigen (W17657) is a novel leukocyte semaphorin-like
 CC protein that stimulates a leukocyte response, including B cell
 CC aggregation, B cell differentiation, B cell survival and/or T cell
 CC proliferation. Its amino acid sequence was derived from a cDNA
 CC clone (T60666) isolated from murine T cells. Human CD100 antigen
 CC (W17657) has also been identified. CD100 polypeptides and fusion
 CC proteins, nucleic acids, and host cells expressing CD100 can be
 CC utilised in diagnostic and therapeutic methods involving modulation
 CC of B and T cell responses, neuron axonal growth and immune cell-
 CC nerve cell interaction.
 SQ Sequence 861 AA;

Query Match 14.4%; Score 606; DB 22; Length 861;
 Best Local Similarity 32.4%; Pred. No. 6,16e-46;
 Matches 143; Conservative 106; Mismatches 145; Indels 47; Gaps 31;

Db 107 eclnyirvlqplsslslyvcgtnafgptcdhnltsfkflgksedgkrcpfdpahsyts 166
 QY 93 DCENITTLER-RSEGLLACGNNARHPSC--WNLVNGTVVPIGEM-RGYAPSPDENSIV 148
 Db 167 vnmvgelysgts-ynfllgsepiisrnsnshspirteyaipwlnepsfvfadvikspdge 225
 QY 149 LFEQDEVYSTIRKQEVNGKIPFRFRIRGESELYTSDTV--MQNPQIKATIVHQ--D--Q 202
 Db 226 geddkvyfftevsveyefvklmiprvarvckgdggllrtliq-kkwtflkarlicskp 284

QY 203 AYDDKIYFFREDNPKNEAPLNVSRVAQLCRGDOGGESSLSVSKWNTFLKAMLVCSDA 262
 Db 285 dsglvfnllqdvflrap-gl-kepvfyavftpqlnnvlsavcaylatveavfgrgky 342
 QY 263 ATKNFNRLQDVFLLPDPGQWRDRTRVGVFSNPWNY---SACVYSLGDIKVF-R-- 315
 Db 343 mqsatveqshtkwryngpvptprpgacidsaaraanytsslnpdktlqfvkdhplmdd 402
 QY 316 --TSSL-----K-G-YHSSLNPRPGKCL-PDQOPTP-TETFOVADRHPE-V-AQR-VE- 360
 Db 403 svtpidnrpklkdvnytgivvdrtqaldgtfydvymfistdrgalghkavilckevh--v 460
 QY 361 PMGPLKT-P-LFHSKVHYQKVAVHRMQASHGETFFHLYLTTRDGTIHKVVPEGEQHSFA 418
 Db 461 --ieetqlfrdsepvltlllsakgrkfyvagsngvvqapiafcekghs-cedcylard 517
 QY 419 FNIMEIQPFERRAAAIQIOTMSLDAER-RKL-YVSSQWEVSQVPLDLCEVYGGGCHGLMSRD 476
 Db 518 pycawspaikacvtlqhceas 538
 QY 477 PYCGWDQG-R-CISIYSSERS 495

RESULT

ID W58540 standard; Protein; 861 AA.
 AC W58540;
 DT 02-SEP-1998 (first entry)
 DE Human semaphorin.
 KW Human semaphorin;
 KW nerve extension inhibiting activity.
 OS Homo sapiens.
 PN J10155490-A.
 PD 16-JUN-1998.
 PR 27-NOV-1996; 332900.
 PA (SUMO) SUMITOMO SEIYAKU KK.
 DR WPI: 98-391044/34.
 DR N-PSDB; V31121.
 PT New human semaphorin gene - useful in the diagnosis of nervous system
 PT and immune disorders
 PS Claim 1: Page 10-12; 15pp; Japanese.
 CC The present sequence represents human semaphorin (translated from the
 CC Japanese specification as semafolin). Semaphorin has nerve extension
 CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
 CC treatment and researches on nervous diseases and immune diseases.
 SQ Sequence 861 AA;

Query Match 14.3%; Score 599; DB 32; Length 861;
 Best Local Similarity 32.4%; Pred. No. 2,99e-45;
 Matches 143; Conservative 105; Mismatches 146; Indels 47; Gaps 31;

Db 107 eclnyirvlqplsslslyvcgtnafgptcdhnltsfkflgksedgkrcpfdpahsyts 166
 QY 93 DCENITTLER-RSEGLLACGNNARHPSC--WNLVNGTVVPIGEM-RGYAPSPDENSIV 148
 Db 167 vnmvgelysgts-ynfllgsepiisrnsnshspirteyaipwlnepsfvfadvikspdge 225
 QY 149 LFEQDEVYSTIRKQEVNGKIPFRFRIRGESELYTSDTV--MQNPQIKATIVHQ--D--Q 202
 Db 226 geddkvyfftevsveyefvklmiprvarvckgdggllrtliq-kkwtflkarlicskp 284
 QY 203 AYDDKIYFFREDNPKNEAPLNVSRVAQLCRGDOGGESSLSVSKWNTFLKAMLVCSDA 262
 Db 285 dsglvfnllqdvflrap-gl-kepvfyavftpqlnnvlsavcaylatveavfgrgky 342
 QY 263 ATKNFNRLQDVFLLPDPGQWRDRTRVGVFSNPWNY---SACVYSLGDIKVF-R-- 315
 Db 343 mqsatveqshtkwryngpvptprpgacidsaaraanytsslnpdktlqfvkdhplmdd 402
 QY 316 --TSSL-----K-G-YHSSLNPRPGKCL-PDQOPTP-TETFOVADRHPE-V-AQR-VE- 360
 Db 403 svtpidnrpklkdvnytgivvdrtqaldgtfydvymfistdrgalghkavilckevh--v 460

Db	225	dgedrvyfftevsveyefvrlvriarvckvgdggllrtlq-kkwtstfkarliscr	283
Qy	202	QAVDDKIYFFREDNPDKNEAPLNVSRVAQLCRDQGGESSLSKWNTEFLKAMLVCSD	361
Db	284	pdsglvfnlrvdfrvrspl-glpvpyfyalftpqldnnvglaavcaynlnstaeefshgk	341
Qy	262	AATNKNFNRLQDVFLLPDPDSGQWRDTRVGVFSNPWNY---SAVCVYSLGDDIKVF---R	315
Db	342	ymgsttveqshktwrvyngvpvxprrgacidsaraanytsslnlnpdktlqfkdphlmd	401
Qy	316	----TS---S-LK-G-YHSLSLNPRGKCL-PQQPPIP-TETFOVADRIPE-V-AQR-VE	360
Db	402	dsvtpidnrprlkkdvnyetqlvvdrtqaldgtvydvfmfstvstdrghalhxai--s-lehav	458
Qy	361	-PMGPKLT-P-LFHSKYHYQKVAVHRMQASHGETHFHVLVLTDRGTIHKVVEPGEQESHF	417
Db	459	hi-leetqlfqdfepvqtllllsskknrfvyagsgnsvvqaplafcgkghgt-cedcvlar	516
Qy	418	AFNIMEIQPFRAAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDLCEVYGGCGHGLMSR	475
Db	517	dpycawspptcatvalhqtspesrgllqemsgdasvcpd-ks-kgsyvrhfkfhgtael	574
Qy	476	DPYCGWDO--GRISIVSSRSRVLSQINP-AEPHKCEPNPKDPKAPLQKVSLAPNSRYVL	532
Db	575	kcsqsknarlvfwkfgngvikaespkyglmrknllif-nlsgdsgvvyqcisee	628
Qy	533	SCPMSRHATYSNRHKNVQSCPEP--GHQSPNCILFIENLTAQYGYHYFCEAQE	585
 RESULT 7			
ID	W51313	standard; Protein; 776 AA.	
AC	W51313;		
DE	08-SEP-1998	(first entry)	
DT	Rat semaphorin W.		
KW	Rat; semaphorin W; nerve extension inhibitor; anti-allergic; anticancer		
KW	Rattus norvegicus;		
OS	Rattus norvegicus.		
PN	WO9815628-A1.		
PD	16-APR-1998.		
PF	03-OCT-1997; J03549.		
PR	09-OCT-1996; JP-287636.		
PA	(SUMU) SUMITOMO PHARM CO LTD.		
PI	Kikuchi K, Kimura T;		
PI	WPI; 98-261015/23.		
PR	N-PSDB; V07279, V07280.		
PT	Nerve extension inhibitor protein semaphorin W - is useful as		
PT	therapeutic drug and diagnostic and research reagent		
PS	Claim 1; Page 60-64; 90pp; Japanese.		
CC	The present sequence represents rat semaphorin W. Semaphorin W and		
CC	its derivatives are nerve extension inhibitors which are useful as		
CC	anti-allergic, immunosuppressant and anticancer agents. The DNA		
CC	encoding semaphorin W can also be used in gene therapy, e.g. using		
CC	a viral vector. The proteins, peptides, DNA and antibodies which		
CC	recognise the protein or peptides, can be used as diagnostic or		
CC	research reagents. Semaphorin W can be used as a screen for		
CC	semaphorin W antagonists with possible therapeutic use.		
CC	Sequence 776 AA;		
Qy	Query Match	12.2%; Score 512; DB 32; Length 776;	
Qy	Best Local Similarity	27.1%; Pred. No. 9,096-37;	
Qy	Matches	152; Conservative 137; Mismatches 218; Indels 53; Gaps	
Db	61	rfaasthynysallvdpashtlyvgardisfaltlpfsggerpridwmpv-ethqncrk	119
Qy	34	RVDFGQTEPTVLHFEGSSVWVGKGYI-L-PDFEGKNASV-RTVNIQSTRGSCU-	89
Db	120	kgkdedchnfigilaivnashlittcgtfakdpkcgvidvssfqverlesgrgkcpfp	179
Qy	90	-DKR-D-CENVITLLERSEGL-LACGTNARHPSCWNL-VNG--TVVPLGEMRGYAPFSP	142
Db	180	aqrsaavmaggvlytatvk-nflgtpeplisravgraedwlrctetlsswlnapafvaamvl	238
Qy	143	DENSLVLFEGDEYVSTRKQYNGKIPRRFRINGESELVYTDV---NONPOFIKATIV	198

```

Db 239 spaewgdegddeiffettetsrvldsyerikvprvarvcagdlgrkltq-grwtftlk 297
QY 199 -H-Q-DQAYDDKIYFFREDPNKPEAPLNVSRVAQLCRGDQGGESSLSVKWNTFLK 254
Db 298 adlcpgeghgrasvlgamaelrpqga-gtpifygiffsqwegaaisavcafrpqdir 356
QY 255 AMLVCSDAATNKNFNRLQVFLPDPGOWRTRVYGVFSPWN--YSAVCVSLGDDID 311
Db 357 avlnpfpelkhdnrglvmndevqprgpcgcinmmklqfsgslpdrvlvlfidrh 416
QY 312 KVFRTS--SLK-GYHSSLP--N---PRPGKLPDQQP--TETQVADRHP--V-AQ 357
Db 417 plmdrpvpadgrpllvtttdtavlrvahrvtslsgkeydvlylgtedghlhravriga 476
QY 358 R-VE-PMGLKT-PLFHSK-YHYQVAVHRMOASGEFHVLYLTTDRGTIHKVVEPGEQ 413
Db 477 -ls-vledial--fpepqpvesmklyhdw--llvgshtevtqvntsngrlgs-csecll 529
QY 414 EHSFAFIMEIQFRRAAAIQMSLDAERRKLYVSQMEVSQVPLDLCEVYGGCHGCLM 473
Db 530 adgpcawsfriacvahagehrmqvdiessadvsclpckepgehpvvfevpatvghvv 589
QY 474 SRDPCGWD-Q-GRGISYSSRSVLQSNPAEPHKECPNPKAPLQKVSLAPNSRY 531
Db 590 lpcspssawascvwhqpsgv 609
QY 532 LSCPMESHRHATYSWRHKENV 551

RESULT 8
ID R71381 standard; Protein; 441 AA.
AC R71381.
DT 21-NOV-1995 (first entry)
DE Vaccinia virus semaphorin IV protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Vaccinia virus.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC ) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131117/17.
DR N-PSDB; Q87443.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 65-67; 101pp; English.
CC The sequence of the vaccinia virus semaphorin IV protein.
CC from variola, based on sequence homology searches of a database with the
CC grasshopper, Tribolium and Drosophila semaphorin sequences. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III
CC (Q87442), vaccinia virus semaphorin IV, Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 441 AA;

Query Match 11.0%; Score 463; DB 13; Length 441;
Best Local Similarity 33.9%; Pred. No. 4.97e-32;
Matches 87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;

Db 115 lvcgtangnpkckw-ldgsddkphgrgyapyqnskvltihsngc-vlsdinisk-eg-i 170

```

```

QY 109 LACGTNARHPSCWNLVNGTVVPLGEMRCYAPFSPDENSLVLEFGDEVYSTRKQYNGKI 168
Db 171 krwrrfdpcgydlytadnvpkdg-lrgafvdkdgtyd-kvyliftdtigr--i-vk 225
QY 169 PRFRIRGE-S-ELYTSDTVMQNPQIKATIVHQDAQYDDKIYFFREDNPKNPEAPLN 226
Db 226 lpytaqmcldneggsslsrhwstflkvelec-dld-grsy-r-q-ihstiktld-nd 279
QY 227 VSRVAQLCRGDOGGESSLSVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLLPDPGOWRD 286
Db 280 tilyvffgspysksalctysmntikqsfstkslegtykqlpsagciclpagkvvphttf 339
QY 287 TRVYGVFSPNPNYSACVYSLGCDIKDVFRTSLKGYHSSLNPNRPCKLCPDQOPIPTTF 346
Db 340 eviekynvlddlikpls 356
QY 347 QVADHRPEVAQRVEPMG 363

RESULT 9
ID R71379 standard; Protein; 730 AA.
AC R71379.
DT 21-NOV-1995 (first entry)
DE Grasshopper semaphorin I protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Grasshopper sp.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC ) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131117/17.
DR N-PSDB; Q87441.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 1; Page 68-72; 101pp; English.
CC The sequence of the grasshopper semaphorin I protein. The proteins
CC encoded by the grasshopper semaphorin I, human semaphorin III (Q87442),
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 730 AA;

Query Match 10.9%; Score 459; DB 13; Length 730;
Best Local Similarity 28.0%; Pred. No. 1.21e-31;
Matches 143; Conservative 130; Mismatches 182; Indels 56; Gaps 46;

Db 32 qfgeervqrfigneshkhfkllkxhdnslivgarnivynislrdltfteqriewhsq 91
QY 29 HVGQDRVD-FQOTEPTVLFH--EPGSSVWVGGRKYLDFDPPEGKN-ASVRTV-NI-G 82
Db 92 ahrelcykgseddcqpyrvlakidddrvlicqtnaykplcrhyalkddgyvvekeye 151
QY 83 STKGSC-L-DRK--DCENYITLLRRSEG--LLACGTNARHPSCWNLVNGT-VVPLG-E 133
Db 152 grglcpgfdpnhstaiyseqlsatv-adfsgtdp-li-yrg-p-lrtersdl-k-qln 204
QY 134 MRGYAPFSPDENSLVLEFGDEVYSTRKQYNGKIYFFREDNPKNPEAPLN 193
Db 205 apnfvn-tmeyndffiffretaveyincgaiky-srvarckhdkgg-phqggdrwtsf 261

```

QY 194 KATIVHODQAYDDKIYYFFREDNPD-KNPEAPLNVSRAQLCRDQGGESSLSVSKWNTF 252
Db 262 lksrlncsvpgdyfyfneiqdstsdiiegnyygg-vekllyvgvfttpvnsiggsavcavs 320
QY 253 LKAMLVCSDAATNK-NFNRLQDVF-LLPDP-SCQWRTRVYGVFSNPWN-Y--SACVYS 306
Db 321 mksllesfdgpfkeqetmnsnwlavpslkvppeprpgqcvndstlpsvsnfvkshlmd 380
QY 307 L-G--D-IDKVFRT--SSLKG-YHS--SL--PNPRPGKCLPDQQPPIPTETFOVADRHEVA 356
Db 381 eaypafft-tpilirlslqyftkiavdqvrtpdgdgkaydvlfigtddgkvikalnsaaf 439
QY 357 QRVPEPMGLPTLPH-S-KYHQKVAV-HRQASHGTFHVLTLTDRGTHHKWVEFG-- 411
Db 440 dssdtvsvvleelqlppgvvknlyvrmddgdsxlvvsvdelaiklhrcgskit 499
QY 412 EOEHSFAFNIME-IQFERRAAAIQTM--SLDAERRKLYVSSQWEVSQVPLDICEVYG-G 466
Db 500 ncrecvslqdpycawdnvelkctavgsdpws 530
QY 467 GCHGCLMSRDPYCGWDQ-G-RCISYSSERS 495

RESULT 10
ID W51314 standard; Protein; 587 AA.
AC W51314;
DT 08-SEP-1998 (first entry)
DE Human semaphorin W
KW Human; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW Immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PN WO9815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR 03-OCT-1996; JP-287636.
PA (SUMO) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-PSDB: V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 10.2%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 1.14e-28;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgddeiyffittetsrafsdyerikvprvarvcagdlggrktlq-qrwttfkadllcp 115
QY 201 DQAYDDKIYYFFREDNPDKNPEAPLNVSRAQLCRDQGGESSLSVSKWNTFLKMLVCS 260
Db 116 gpehgrassvlqdvavlrpelga-gtgifgissqvegatisavcafrpqdirtvlnp 174
QY 261 DAATNKNFNRLQDVLPLDPDSQGWDRTRVYGVFSNPWN---YSACVYSIGDIDKVFRTS 317
Db 175 frelkhdnrglpvndvpqprgpcitnmklrhfgsslsipdrvltrfirdhplmdrp 234
QY 318 --SLK-GYHSSLP---N---PRPGKCLPDQQPPI--TETFOVADRHE-V-AQR-VE-P 361
Db 235 vfpadghpllvtttdtaylorvahrvtlskeydvlylgtedghlhravrigaq-ls-vl 292
QY 362 MGPUKT-PLFHSK-YHQKVAHMQASHGETHFVLTLTDRGTHHKWVEFGSOEHSFAF 419
Db 293 edial--fpepgvcmnkl-yhsw-llvgstevtqvnttcnrglqs-csecilagdpvc 347

QY 420 NIMEIQFERRAAAIQTMSLDAERRKLYVSSQWEVSQVPLDLCEVYGGCHGCLMSRDPYC 479
Db 348 awefrldecvahagehrglvqdiesadvsllcpkegepvpvfevpyataahvllscps 407
QY 480 GWD-Q-GRCISYSSERSVLQSNPAEPHKECPNPKAPDKVSLAPNSRYLLSCPME 537
Db 408 sawascvwhqpsgv 421
QY 538 SRHATYSWRHKENV 551
RESULT 11
ID W64221 standard; Protein; 974 AA.
AC W64221;
DT 06-OCT-1998 (first entry)
DE Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN WO9827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GEM) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallee ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-362774/31.
DR N-PSDB: V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17j; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 9.3%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 4.83e-25;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechmfikvllkndalifvcetnafnpscrnykndtlepfdfgdfsgmarcpdydahan 172
QY 91 KRDCENVIT-LLERRSEGLACGTNARHPSCWNLVNGTVVPLG-EMRGVA--PFSPDENS 146
Db 173 valfadgklysatvt-dflaidavlyrs1-gesp--tlrtvkhdkwlepyfvqadvy 228
QY 147 LVLFEGDEVYSTRKOEYNG-KIPFRRIRGESELYTSDTMQNPQFIKATIVHQDQAYD 205
Db 229 dviylffireiaeyntmkgvvpvraqvckndmggsqrveleqwtstflkarlncsvpgds 288
QY 206 DKIIYFFRGDNPDKNPEAPLNVSRAQLCRDQGGESSLSVSKWNTFLKAMLVCSDAATN 265
Db 289 hfynilqavtdviring--rdv-vlatfstpynsipsavcaydmdlasvftgrfkeq 345
QY 266 K-NFNRLQDVLPLDPDSQGWDRTRVYGVFSNPWN-Y--SACVYSIGDIDKVP--RTSSL 319
Db 346 kspdstwtvpdervkprpgccagsssleryatsnefddtlnfkthplmdaevpsif 405
QY 320 KGYHSS---LPNPRPGKCLPD--QQPIPTETFOVADRHEVA-Q--RVEP-M--G-P-L- 365

```
Db 406 nrpvflrtmrvrltkiavdaagpy-qnhtvrvflgsekgiilkflaigsgflndself 464
Qy 366 KTLPLF-HS--KYHOKVAVHRMQASHGHTFVLYLTTRGTIHKVVEP-GSQE--HSPAF 419
Db 465 leensvynsekcsydgvedkrimglqdrassslyvafstcvikvplgrcerhkg-ckkt 523
Qy 420 -NIMEI-QPERRA-AAIQ--TMSLDAER--RKLYVSSQWESQVPLDLCEVYGGGCH-G 470
Db 524 ciasrdpvcywikeggacshlpsnrlfcdiergn 560
Qy 471 CLMSRDPYCGW--DQGRCSISYSSERSVL-QSINPAE 504

RESULT 12
ID R71384 standard; Protein; 712 AA.
AC R71384;
DE 21-NOV-1995 (first entry)
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Tribolium sp.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC ) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB; 087446.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 85-88; 101pp; English.
CC The sequence of the beetle Tribolium semaphorin I protein. The gene was
CC isolated by PCR using Tribolium genomic DNA. The proteins encoded by the
CC grasshopper semaphorin I (Q87441), human semaphorin III (Q87442),
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I or variola major (smallpox) virus
CC semaphorin IV (Q87447) genes were used to generate a series of peptides
CC (R70370-R70418), which retain semaphorin receptor binding activity. The
CC semaphorin derived or semaphorin receptor derived peptides are potent
CC modulators of nerve cell growth, immune responsiveness and viral
CC pathogenesis. They can be used in diagnosis and treatment of neurological
CC disease and neuro-regeneration, immune modulation and diagnosis and
CC treatment of viral and oncological infection and diseases.
SQ Sequence 712 AA;

Query Match 9.0%; Score 378; DB 13; Length 712;
Best Local Similarity 27.0%; Pred. No. 6.65e-24;
Matches 140; Conservative 128; Mismatches 194; Indels 56; Gaps 43;

Db 48 hfivlvg-detsilvggrnrvynisifalserkgrgidwps-sdahgqlcilkgtdddc 105
Qy 43 HTVLFEPCSSVWVGGRKYI-L--FDPEGKNASVTVNIGSTKGS-C-L-DKRD--C 94

Db 106 qnyirilysepgklivictqnsyplcrtayfkegkylvekevegigicpynpohnstsv 165
Qy 95 ENYITLLRRRSEG-LLACGTNARHPSCNVLV--NGT-VVPLG-ENRGYAPFSPDENSL-V 148

Db 166 syng-qlfsatv-adfsggdpllyrepartel-sdlkqlnapfnvs--v----aygdyi 216
Qy 149 LFEGDEVYSTIRKQBYNGKIPFRFRIRGESELYTSDTMQNPNQFIKATIVHQDQAYDDKI 208

Db 217 fffretaveymcgkviy-srvarvckddkgy-phqdrdrwtsflkarlncsipgeyfp 274
Qy 209 YFFREDNPD-KNPEAPLVNVRVAQLCRDGGGESSLSVSKWNTFLKAMLVCSDAATNK- 266

Db 275 yfdeiqstdivregrynsddskkilylltppvnaiggsaicaaymadilrvfegsfkhq 334
Qy 267 NFNLQDVFLLPDPSSGQWRDR--VYGVFSNPNW-Y--SACVYSLGIDIDKVF----R 316
```

```
Db 335 etinsnlpvppqnlvpeprgqcvdrsilpdknvnfikkthslmed-vpalfg-kpvlvr 392
Qy 317 SSLKG-Y-H-S-SL-PNPRPGKCLPDQOPIPTETFOVADRHPEVAQRVPEPMGLKTLPLFH 371
Db 393 vslgyrftaivdpqvktnnqyldvlygtdddkvfkavnpkrhakallyrkyrtsvh 452
Qy 372 -S-KYHKQKVAHV-RMQASHGSETFHVLYLTTRGTIHKVVE-PGEQEHSAFNTMEIQPF 427
Db 453 phgapvklkiapgygkvvvvgkdeirlanlnhcas-ktrckdcvelqdpchcavdaknl 511
Qy 428 RRAAAIQTMSLDAERRKLYVSSQWESQVPLDLCEVYGGGCHGLMSRDPYCGWD--QGR 485
Db 512 cvsldtvsyrlfllqdvvgddnk-cwspqtdckktvik 548
Qy 486 CISIYS-SE-RSVLOSINPAEPHKECPNPKDPKAPLQK 521

RESULT 13
ID R71382 standard; Protein; 650 AA.
AC R71382;
DE 21-NOV-1995 (first entry)
DE Drosophila semaphorin I protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Drosophila sp.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC ) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR N-PSDB; 087444.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 74-77; 101pp; English.
CC The sequence of the Drosophila semaphorin I protein. The gene was
CC isolated by PCR using primers based on sequence homology between the
CC complete grasshopper and partial Tribolium and Manduca (moth) semaphorin
CC gene sequences. The products gave two different sequences, each of which
CC was used to obtain its respective complete sequence: semaphorin I and II
CC (Q87445). The proteins encoded by the grasshopper semaphorin I (Q87441),
CC human semaphorin III (Q87442), vaccinia virus semaphorin IV, Drosophila
CC semaphorin I and II, Tribolium semaphorin I (Q87446) or variola major
CC (smallpox) virus semaphorin IV (Q87447) genes were used to generate a
CC series of peptides (R70370-R70418), which retain semaphorin receptor
CC binding activity. The semaphorin derived or semaphorin receptor derived
CC peptides are potent modulators of nerve cell growth, immune
CC responsiveness and viral pathogenesis. They can be used in diagnosis and
CC treatment of neurological disease and neuro-regeneration, immune
CC modulation and diagnosis and treatment of viral and oncological infection
CC and diseases.
SQ Sequence 650 AA;

Query Match 8.7%; Score 367; DB 13; Length 650;
Best Local Similarity 28.4%; Pred. No. 7.32e-23;
Matches 124; Conservative 104; Mismatches 158; Indels 51; Gaps 38;

Db 3 dcqnyirilmvpsgrlfcvctnsfrpmcntyilsdsnytleatkngavcvcprhnst 62
Qy 93 DCENYITLLRRRSEG-LLACGTNARHPSCWN-LYNGTVVPLGEMR-GYA--PFPDENSL 147

Db 63 svladnelysgtv-adfsgsdpllyreplqtceydsils-lnapnfv-ssft-qgd-f--- 114
Qy 148 VLFEDEVYSTIRKQBYNGKIPFRFRIRGESELYTSDTMQNPNQFIKATIVHQDQAYDDK 207

Db 115 yffretavefincgkaiy-srvarvckdkgghfr-nrwtsfklrncisipgdy 172
Qy 208 IYFFREDNPD-KNPEAPLVNVRVAQLCRDGGGESSLSVSKWNTFLKAMLVCSDAATNK 266
```


QY 471 CLMSRDPYCGWDQGR-CISIIYSSE-RSVLQSIINPAE-PHKECPN 511

Search completed: Thu Jul 8 19:51:43 1999
Job time : 38 secs.

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:50:13 1999; MasPar time 25.88 Seconds
873.182 Million cell updates/sec
Tabular output not generated.

```
>US-09-041-236-2
Title:
Description:
Perfection Score: 4201
Sequence: 1 AVVKGHVGDYDFDGGTEPH.....IENLTAQ
(24-587) from US09041236.pep (41 of 45)
```

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:
pir60
  1:pir1 2:pir2 3:pir3 4:pir4
```

Statistics: Mean 48.923; Variance 85.907; scale 0.569

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query		Length	DB	ID	Description	Pred. No.
	Score	Match					
1	1850	44.0	653	2	T03102	semaphorin homolog A3	0.00e+00
2	737	17.5	748	2	I48744	semaphorin A - mouse	2.86e-133
3	725	17.3	751	2	I48748	semaphorin E - mouse	1.71e-133
4	716	17.0	749	2	G01856	semaphorin V - human	2.06e-133
5	688	16.4	753	2	G02173	semaphorin III family	5.96e-122
6	681	16.2	772	2	A49059	collapsin - chicken	2.44e-122
7	669	15.9	772	2	I48747	semaphorin D - mouse	1.41e-111
8	664	15.8	666	2	I58169	semaphorin III - mouse	1.98e-111
9	656	15.6	771	2	D49423	semaphorin III precur	1.37e-111
10	618	14.7	782	2	I48746	semaphorin C - mouse	6.91e-108
11	522	12.4	834	2	S66498	M-sena F protein prec	4.17e-86
12	474	11.3	760	2	I48745	semaphorin B - mouse	2.35e-75
13	470	11.2	403	2	E42521	A39R protein - vacci	1.83e-74
14	463	11.0	441	2	S29921	hypothetical protein	6.59e-73
15	461	11.0	730	2	JH0798	hyctactin IV precurs	1.83e-72
16	448	10.7	1074	2	JC5928	semaphorin F precurs	1.40e-69
17	389	9.3	711	2	A49423	semaphorin I precurs	1.26e-56
18	367	8.7	656	2	B49423	semaphorin I - fruit	7.32e-52
19	362	8.6	724	2	C49423	semaphorin II precurs	8.75e-51
20	293	7.0	295	2	J01775	SalligR protein - vacc	3.60e-36
21	195	4.7	1884	2	JC4975	plexin 2 precursor -	7.92e-17
22	171	4.1	1905	2	I51553	Plexin - African claw	3.24e-12
23	170	4.0	1872	2	JC4976	Plexin 3 precursor -	4.90e-12


```
Qy 304 VYSLGDIKVERTS-SLK-G--YH--S--LPNPRGKCLPDQ-QPI-PTETF--QV-A 349
Db 402 FARSHPAMNVPFINS-RPIMIKTDVXYOFTQIVVDRVDAEDGQ-YDVMFIGFDIGTVL 459
Qy 350 -DR-HPEVAQVPEMGLKPLFLHSK--YHYQKVAVHRMQASHGETHFVLYLTTRDTGTH 405
Db 460 KWSIPKRETHLEBEVLEEMTVPRETVISAMKISTKOOLYIGSATGVSQPLHRCDV 519
Qy 406 KVVE-PGEQHSFAFIME-IQPFRAAAIQTMSLDAERKLYVSSQWESQVPLDLCEV 463
Db 520 YGKACACCLLARDPYCWDGSSCSRYFPTAKRRTRDINGDPLTHCSDLQHDHNSGQ 579
Qy 464 YGGCHGCLMSRDPYCGWDGRCISYSS-RSVL-QSINPAEPHKECPN-PKPDKAP-- 518
Db 580 TLEEKIIVGVNSTFLECSPKSORAIVY-WQFQKQNDHKE 621
Qy 519 -LQ-KVSLA-PNSRYILSC-PMESRHATYSWR-HKENVEQSC 556

RESULT 7
ENTRY semaphorin D - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM #formal_name Mus musculus #sequence_revision 02-Jul-1996 #text_change
DATE 02-Jul-1996
ACCESSIONS I48747
REFERENCE I48747
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48747
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-772 #label RES
#cross-references EMBL:X85993; NID:g854329; PID:g854330
GENETICS
#gene semD
#superfamily semaphorin
CLASSIFICATION #length 772 #molecular-weight 88710 #checksum 1776
SUMMARY

Query Match 15.9%; Score 669; DB 2; Length 772;
Best Local Similarity 30.2%; Pred. No. 1.41e-119;
Matches 170; Conservative 143; Mismatches 182; Indels 67; Gaps 51;

Db 58 HTFLLDE-ERSRLVYGAKDHFISNLNIXDKFQIVVPVY-TRRDCKWAGKILKECA 115
Qy 43 HTVLFEPGSSVWVGKGYLDFPEGKN-AS-VRTNIGSTKGSC--LDK---RDCE 95
Db 116 NFIKVLAYNOTHLYACGTGAHFPICTYIEVGHHPEDNIFKLQDSHFENGRGKSPYDKL 175
Qy 96 NYITLLERSEG-LLACGTNARHPSCNWL-V--N--GTVVPL--G--EM-RGAFPSFDE 144
Db 176 LTASLLIDGELYSGTA-ADFMGRDFAIFRTLGHHPIRTQHDNRWLNDRPFIASHIPE 234
Qy 145 NSLVFEGDEVYSTIRKQYNGK-IPFRIRIGSELYTS-DIV-MQNQFIKATIVHQ 200
Db 235 SDNPEDKVVFFFRFENAIAGEHSGKATHAIGQICKNDFGHRSL-VNKKWTFLLKALBIC 293
Qy 201 -DOAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGDGGSLSVSKNTFLKMLVC 259
Db 294 SVPCNGIDTHFDELQDVELM-NSKDP-KNPIVYGVFTTSSNIFKGSAYCMYSNDRVY 351
Qy 260 S--DA-ATNKNFNRLQDVLDPDPGQWRDTRVYGVFSNPNW-Y--SAVCYISLGDIDKY 313
Db 352 FLAGPAHRDGPNYQWVPYQGRVPYPRPGTCSPTFGGDFDSTKDLDDVITFGRSHPAMYN 411
Qy 314 FRIS-SLK-G--YH--S--LPNPRGKCLPDQ-QPI-PTETF--QV-A-DR-HPEVAQ 357
Db 412 PVFPINN-RPIMIKTDVNYOFTQIVVDRVDAEDGQ-YDVMFIGFDIGTVLKVVSVPKETW 469
Qy 358 RVEPMGLKPLFLHS--KYHYQKVAVHRMQASHGETHFVLYLTTRDTGTHKVE-PGEQE 414
```

```
Db 470 HDLEEVLLEEMTVFERPTTISAMELSTKQOOLYIGSTAGVAQLPHRCDIYKACAECC 529
Qy 415 HSFAFIME-IQPFRAAAIQTMSLDAERKLYVSSQWESQVPLDLCEVYGGGCHCLM 473
Db 530 ARDPYCAMWDGSSCSRYFPTAKRRTRDINGDPLTHCSDLQHDHNSHPSLEERIYGV 589
Qy 474 SRDPYCGWDGRCISYSS-RSVL-QSINPAEPHKECPN-PKPDK--AP-LQ-KVSLA- 525
Db 590 ENSSTFLECSPKSORALVY-WQ 610
Qy 526 PNSRYILSC-PMESRHATYSWR 546

RESULT 8
ENTRY semaphorin III - mouse (fragment)
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM #formal_name Mus musculus #sequence_revision 26-Jul-1996 #text_change
DATE 04-Sep-1998
ACCESSIONS I58169
REFERENCE I58169
#authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;
Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#journal Neuron (1995) 14:949-959
#title Semaphorin III can function as a selective chemorepellent to
pattern sensory projections in the spinal cord.
#cross-references MUID:95267432
#accession I58169
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-666 #label RES
#cross-references GB:L40484; NID:g703189; PID:g703190
GENETICS
#gene SemIII
#superfamily semaphorin
CLASSIFICATION #length 666 #checksum 9654
SUMMARY

Query Match 15.8%; Score 664; DB 2; Length 666;
Best Local Similarity 30.3%; Pred. No. 1.98e-118;
Matches 150; Conservative 128; Mismatches 167; Indels 50; Gaps 39;

Db 60 RGKSPYDPKLLTASLLIDGELYSGTA-ANFMGRDFAIFRTLGHHPHPIRTQHDNRWLN 118
Qy 135 RGVPSPDENSLVLEEGDEVYSTIRKQYNGK-IPFRIRIGSELYTS-DIV-MQNP 190
Db 119 RFTSAHLIPSDNPEDKVVFFFRFENAIAGEHSGKATHAIGQICKNDFGHRSL-VNKK 177
Qy 191 QFIKATIVHQ-DOAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGDGGSLSVSKW 249
Db 178 TTELKARLICSVPGPNGIDTHFDELQDVELM-NSKDP-KNPIVYGVFTTSSNIFKGSAYC 235
Qy 250 NTFKLAWLVCS--DA-ATNKNFNRLQDVLDPDPGQWRDTRVYGVFSNPNW-Y--SAVC 303
Db 236 MYSMSDVRRVLLGPYAHRDGPNTQWVPYQGRVPYPRPGTCSPTFGGDFDSTKDLDDVIT 295
Qy 304 VYSLGDIKV---F--RTSS-LK--GYHSSLNPNRPGKCLPDQ-QPI-PTETF--QV-A 349
Db 296 FARSHPAMNVPFINS-RPIMIKTDVNYOFTQIVVDRVDAEDGQ-YDVMFIGFDIGTVL 353
Qy 350 -DR-HPEVAQVPEMGLKPLFLHS--KYHYQKVAVHRMQASHGETHFVLYLTTRDTGTH 405
Db 354 KVVSVPEKETHLEBEVLEEMTVPRETTISAMELSTKQOOLYIGSTAGVAQLPHRCDI 413
Qy 406 KVVVE-PGEQHSFAFIME-IQPFRAAAIQTMSLDAERKLYVSSQWESQVPLDLCEV 463
Db 414 YGKACACCLLARDPYCWDGSSCSRYFPTAKRRTRDINGDPLTHCSDLQHDHNSHGP 473
Qy 464 YGGCHGCLMSRDPYCGWDGRCISYSS-RSVL-QSINPAEPHKECPN-PKPDK--AP 518
Db 474 SLBERIIVGVNSTFLECSPKSORALVYVQFORNEDRKEETKMGDHIIRTSQGLLRS 533
Qy 519 -LQ-KVSLA-PNSRYILSC-PMESRHATY-SW-RHKENVEQSCPEG-HQ-SPNCILFIEN 570
```

REFERENCE	I48744
#authors	Puschel, A.W.; Adams, R.H.; Betz, H.
#journal	Neuron (1995) 14:941-948
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.
#cross-references	MUID:95267431
#accession	I48746
#status	preliminary; translated from GB/EMBL/DDBJ
#molecule_type	mRNA
#residues	1-782 #label RES
#cross-references	EMBL:X85992; NID:g854327; PID:g854328
GENETICS	
#gene	semc
SUMMARY	#length 782 #checksum 1571
Query Match	14.7%; Score 618; DB 2: Length 782;
Best Local Similarity	33.7%; Pred.No. 6,91e-108;
Matches	149; Conservative 96; Mismatches 148; Indels 49; Gaps 32;
Db	74 KRCQNYKILLPLNSHLLTCGTAAPSLCAIYIHIASTFLAQDEAGNVILEDKGKHGCPF 133 : : : : : :
Qy	91 KRDCENYIT-LLERRSEGLLACGTNARHPSC-W-N-----LV-NCTV-VPLGEMRGYAPF 140 : : : : : : : :
Db	134 DPNEKSTALVVDELTYGTVS-SFGONDPAISRSOSSRPTKTTESSLNLQDPAPFVASATS 192 : : : : : : : : : :
Qy	141 SPDENSLVLFEGDEVYSTIRKQYNGKIPFRFRIERGESELYTSDTV--MQNPQFIKATIV 198 : : : : : : : : : :
Db	193 PESLGSPIGDDDKYIIFFSETGOEFFFFENTI-VSRVARVKCGDEGGERVLO-QRWTSFL 250 : : : : : : : : :
Qy	199 HQDO-A-Y--DDKIYYFFREDNDP-KNPEAPLNVRQAOLCRDGGESSUSVKWNFTL 253 : : : : : : : : :
Db	251 KAQLLCRPDDGPFNFVLQDVFTLPNPO-DWRKTLISGVFTSQWHRGTTEGSALICVTM 309 : : : : : :
Qy	254 KAMLVCSDAATKNFNRLQDVELL-PDPSQWRDFTVYGFSPNN---Y--SAVCYVSL 307 : : : : : :
Db	310 NDQVKAPDGLYKKVNRETQWTTTHQVTPPRGACITNSAREKINSLSLOLPDRVNLFL 369 : : : : : : : :
Qy	308 GDIDKVP-----RTSS-LKG-Y-HSS-LNPNRPGRCLPD--QQPIPTTFQVADRHPEVA 356 : : : : : :
Db	370 KDHFMDGQVRSLLLLQPAPRVAVHRVPGCHS-IYDLVFLCTGDGRLHKAVTLSSR 428 : : : : : :
Qy	357 QRVEPM-GPLTKPLF--HSKYHQKVAVHRMOASHGETFHVLYLTDTDRGTHKVVPEGEQ 413 : : : : : :
Db	429 VH---I-IEELQIFPOGPQNLLDSHGGLLYASSHSGVQPVANCSLYPT-CGCDLL 483 : : : : : : : : :
Qy	414 EHSEAFNIMEIQFRAAAIQMTSLDAERKLKLYSSQWEVSQVPLDLCEVGGGCHGCLM 473 : : : : : : : : :
Db	484 ARDPYCWTSGACRLASLYOPD 505 : : : : : : : : :
Qy	474 SRDPCGWDOGRC-I-SIYSSE 493 : : : : : :
RESULT	11
ENTRY	S66498 #type complete
TITLE	M-sema F protein precursor - mouse
ORGANISM	#formal_name Mus musculus #common_name house mouse
DATE	28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 10-Sep-1997
ACCESSIONS	S66498
REFERENCE	S66498
#authors	Inagaki, S.; Furuyama, T.; Iwahashi, Y.
#journal	FEBs Lett. (1995) 370:269-272
#title	Identification of a member of mouse semaphorin family.
#cross-references	MUID:95385809
#accession	S66498
#status	preliminary
#molecule_type	mRNA
#residues	1-834 #label INA
#cross-references	EMBL:S79463; NID:g1110598; PID:g1110599
FEATURE	
-1-21	#domain signal sequence #status predicted #label SIG\
22-834	#product M-sema F protein #status predicted #label MAT
SUMMARY	#length 834 #molecule_weight 92556 #checksum 7199

Query Match	12.4%;	Score 522;	DB 2;	Length 834;
Best Local Similarity	28.5%;	Pred. No.4.17e-86;		
Matches	142;	Conservative	118;	Mismatches 190; Indels 49; Gaps 36;
Db	102	KGKS-NOTECNFIRPLOPYNSSHLYVCGTYAFQPKCYIYNWLFTLDRAFEDGKGKCP	160	
Qy	85	KGCLDKRDCENYITLLER-RSEGLLAGCTNARHPSC-W-NLVNGTV--VPLGEMRGYAP	139	
Db	161	YDPAKGHTGLLVGDEYSATLN-NFLGTEPVLIRYMGTHHSIKTEYLAFLWNEPHFVGS	219	
Qy	140	FSPDENSLVLFEGDEVYSIRKQEYNGKIPRRERRIGESELYTSD--TV-MQNQFIKAT	196	
Db	220	FYPESVGSTGDDDKTYFFFSERAVEYDCYSEQVVARVAVCKGDMGAGTILQ-KKWTTF	278	
Qy	197	IVHOD-QAY--DDKTYFFREDNPDKNPEAPLNVSRVAQLCRDQGOGESSLSYKWNTF	252	
Db	279	LKARLVCSAPDKMKVFNQLKAVHTLUGAS--WHNTTFEGVQARWDMDLNAVCEYOLEQ	336	
Qy	253	LKAMLVCSDAATNKNFRLOQVFLLPDPSGQWRDTRVIGVFSNPW---NYSAVCVYSLGD	309	
Db	337	IQOVFEGPYKEYSEQAKWARYTDVPSPRGSCINNHRONGYTSLSLELPDNTLNFKK	396	
Qy	310	IDKVF---RT-S-SLK-G-YHSLPLNPRCKCLPD-QQPIP-TETFOVADRHE-VA-	356	
Db	397	HPLMEDQVAPRLGRPLLVKNTNFTHVADVDPGLDGYATVYVLTGTDGMLLKAVSLGP	456	
Qy	357	QR-VE-PMGP-LKTPLFHSKY-HYQKVAVHRMQASHGETFHVLYLTDRGTIHKWVEPGE	412	
Db	457	WTH--M-VEELQVDOEP-VESLVLSQSKVLFAGRSOLVOLSLADCTKYRF-CVDCV	510	
Qy	413	QEHSAFATNWEQPFRAAAIOTMSLDERRKLYVSSQEWESQVPLDCEYVGGGCHGL	472	
Db	511	LARDPYCAWNAVTSRCVATTSGRSFVLQHVANLDTSKMCNQYIKKVRSPKPNITVVS	570	
Qy	473	MSRDPYCGND-Q-GRCISIIYSER-SVL-QSINPAEPKHEKCPNPKPDKA-PLQK-VSLAP	526	
Db	571	GTDLVLPCHLSNLAHAHW	589	
Qy	527	NSRYLSCPMESRHATYSW	545	
RESULT	12			
ENTRY	I48745	#type complete		
TITLE	semaphorin B - mouse			
ORGANISM	#formal_name Mus musculus	#common_name house mouse		
DATE	02-Jul-1996	#sequence_revision 02-Jul-1996	#text_change 28-Feb-1997	
ACCESSIONS	I48745			
REFERENCE	I48744			
#authors	Puschel, A.W.; Adams, R.H.; Betz, H.			
#journal	Neuron (1995); 14:941-948			
#title	Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension.			
#cross-references	MUID:95267431			
#accession	I48745			
#status	preliminary; translated from GB/EMBL/DBJ			
#molecule_type	mRNA			
##residues	1-760	##label RES		
GENETICS	##cross-references EMBL:X85991; NID:g8543325; PID:g8543326			
#geno	semb			
SUMMARY	#length 760	#molecular-weight 83458	#checksum 2188	
Query Match	11.3%;	Score 474;	DB 2;	Length 760;
Best Local Similarity	30.2%;	Pred. No.2.35e-75;		
Matches	150;	Conservative	111;	Mismatches 183; Indels 52; Gaps 37;
Db	123	QCENFTRVLVSYNATHLYACGTFAFSPACTFTLEQDSLLLPILIDKVDGKGQSPILFT	182	
Qy	93	DCENYITLLER-RSEGLLAGCTNARHPSC-W-NLVNGTVTP-LGE--M--RGYAPFSP-D	143	
Db	183	STQAVLVDG-MLYSGT-MNNFLGSEPIMLRTLGSHPLKTDIFLRWLH-ADAFSVAIPTS	239	

RESULT	14	
ENTRY	S2921	#type complete
TITLE	hypothetical protein 15 - vaccinia virus	
ORGANISM	#formal_name vaccinia virus	
DATE	20-Feb-1995	#sequence_revision 20-Feb-1995
	09-Sep-1997	#text_change
ACCESSIONS	S2921	
REFERENCE	S2907	
#authors	Amegadzie, B.Y.	
#submission	submitted to the EMBL Data Library, January 1991	
#accession	S2921	
#status	preliminary	
#molecule_type	DNA	
#residues	1-441	#label AME
#cross-references	EMBL:X57318; NID:g62239; PID:g62254	
SUMMARY	#length 441	#molecular_weight 50185
		#checksum 6034
Query Match	11.0%	Score 463; DB 2: Length 441;
Best Local Similarity	33.9%	Pred. No. 6.59e-73;
Matches	87; Conservative	57; Mismatches 96; Indels 17; Gaps 16;
Db	115	LVCGTNNGPKWK - IDGSDDPKRRGRGYAPXONSKVTIISHNGC-VLSDINISK-EG-I 170
QY	109	LACGTNARHPSCNLWNVLPGLGEMRGYAPFSPDENSELVLFEGDEVYSTRKQYNGKI 168
Db	171	KWRREDGCGDYLDYADNVIPKDG-LRGAFYDKDGTVD-KVYLFTDTIGSKR--I-VK 225
QY	169	PRFRIRGE-S-ELYTSDTVMONPOFIKATIVHODQAYDDKIIYFFREDNPKNPEAPLN 226
Db	226	IPYIAQMLNDEGGPSLSHRWSFVKVELEC-DID-GRSY-R-Q-IHSRTIKTD-ND 279
QY	227	VSRAQLCRDQGESSLSVSKWNTFLKAMLVCSDAANKFNRLQDVFLLPDPSSGQWRD 286
Db	280	TILYFEDSPYSKALCTYMTNTIKQSFSTSKLEGYTKQLPSPASGICLPAGKVVPHHTF 339
QY	287	TRYGVESPNWYSAVCYVSLGIDKVFRTSLKGYHSSLNPRGKCLPDQPIPTETF 346
Db	340	EVIEKYNVLDDIIKPLS 356
QY	347	QVADRHPEVAQRVEPMG 363
RESULT	15	
ENTRY	JH0798	#type complete
TITLE	fascilin IV precursor - American bird grasshopper	
ORGANISM	#formal_name Schistocerca americana	#common_name American bird grasshopper
DATE	30-Sep-1993	#sequence_revision 30-Sep-1993
	09-Sep-1997	#text_change
ACCESSIONS	JH0798	
REFERENCE	JH0798	
#authors	Kolodkin, A.; Matthes, D.J.; O'Connor, T.P.; Patel, N.H.; Admon, A.; Bentley, D.; Goodman, C.S.	
#journal	Neuron (1992) 9:831-845	
#title	Fascilin IV: Sequence, expression, and function during growth cone guidance in the grasshopper embryo.	
#accession	JH0798	
#molecule_type	mRNA	
#residues	1-730	#label KOL
#cross-references	GB:L00709; NID:g160844; PID:g160845	
#experimental_source	embryo	
COMMENT	This protein plays a role in growth cone guidance in the developing central nervous system.	
KEYWORDS	glycoprotein; transmembrane protein	
FEATURE		
1-22	#domain signal sequence	#status predicted
23-730	#product fascilin IV	#status predicted
23-627	#domain extracellular	#status predicted
628-652	#domain transmembrane	#status predicted
653-730	#domain intracellular	#status predicted
471,163,267,360,539	#binding_site carbohydrate (Asn) (covalent)	#status

THIS PAGE BLANK (USPTO)

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-waterman algorithm

Run on: Thu Jul 8 19:48:22 1999; MasPar time 17.96 Seconds
Tabular output not generated. 887.732 Million cell updates/sec

Title: >US-09-041-236-2
Description: (24-587) from US09041236.pap (41 of 45)
Perfect Score: 4201
Sequence: 1 AWWKHVGQDRVDFGQTEPH.....IENLTAQQYGHVFCERAOEGS 564

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 50.030; Variance 75.658; scale 0.661

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	470	11.2	1 VA39_VACCC	PROTEIN A39.	2.42e-86
2	463	11.0	1 VA39_VACCV	PROTEIN A39.	1.53e-84
3	163	3.9	1 SEX_HUMAN	TRANSMEMBRANE PROTEIN	7.89e-13
4	105	2.5	1 U508_HCMVA	HYPOTHETICAL PROTEIN H	3.43e-02
5	102	2.4	1 NRT2_RAT	T-CELL ECTO-ADP-RIBOSY	1.01e-01
6	99	2.4	1 NRT1_RAT	T-CELL ECTO-ADP-RIBOSY	2.88e-01
7	99	2.4	1 EFLA_TRIRE	ELONGATION FACTOR 1-AL	2.88e-01
8	99	2.4	1 VRP2_SALTY	65 KD VIRULENCE PROTEI	2.88e-01
9	99	2.4	1 VRP2_SALCH	65 KD VIRULENCE PROTEI	2.88e-01
10	99	2.4	1 VRP2_SALEN	65 KD VIRULENCE PROTEI	2.88e-01
11	99	2.4	1 VRP2_SALDU	65 KD VIRULENCE PROTEI	2.88e-01
12	99	2.4	1 PMA1_AJECA	PLASMA MEMBRANE ATPASE	2.88e-01
13	95	2.3	1 HA21_HUMAN	HLA CLASS II HISTOCOMP	1.12e+00
14	95	2.3	1 VL2_HPV5B	MINOR CAPSID PROTEIN L	1.12e+00
15	95	2.3	1 VL2_HPV05	MINOR CAPSID PROTEIN L	1.12e+00
16	95	2.3	1 AMOH_ARTGO	HISTAMINE OXIDASE (EC	1.12e+00
17	98	2.3	1 DNB1_HSV6U	MAJOR DNA-BINDING PROT	4.06e-01
18	95	2.3	1 MET_HUMAN	HEPATOCYTE GROWTH FACT	1.12e+00
19	94	2.2	1 RSR_ECOLI	RIBOSE OPERON REPRESSO	1.56e+00
20	91	2.2	1 DDL_ENTFA	D-ALANINE--D-ALANINE L	4.14e+00
21	91	2.2	1 VP39_NPVLD	MAJOR CAPSID PROTEIN.	4.14e+00
22	93	2.2	1 DP3B_MYCPN	DNA POLYMERASE III, BE	2.17e+00
23	92	2.2	1 CEM1_YEAST	3-OXOACYL-[ACYL-CARRIE	3.00e+00

24	92	2.2	1 EFLA_NEUCR	ELONGATION FACTOR 1-AL	3.00e+00
25	91	2.2	1 ELAS_PSEAE	PSEUDOLYSIN PRECURSOR	4.14e+00
26	91	2.2	1 NOT_HUMAN	IMMEDIATE-EARLY RESPON	4.14e+00
27	91	2.2	1 GCVK_HCMVA	GANCICLOVIR KINASE (EC	4.14e+00
28	93	2.2	1 Y223_METJA	HYPOTHETICAL PROTEIN M	2.17e+00
29	94	2.2	1 PMA2_YEAST	PLASMA MEMBRANE ATPASE	1.56e+00
30	94	2.2	1 N100_YEAST	NUCLEOPORIN NUP100/NSP	1.56e+00
31	94	2.2	1 POL_HTLV2	POLYPROTEIN (CONTA	1.56e+00
32	93	2.2	1 ENDI_YEAST	VACUOLAR BIOGENESIS PR	2.17e+00
33	94	2.2	1 ENTK_MOUSE	ENTEROPEPTIDASE (EC 3.	1.56e+00
34	94	2.2	1 IREL_YEAST	SERINE/THREONINE-PROTE	1.56e+00
35	92	2.2	1 DNB1_HSV6Z	MAJOR DNA-BINDING PROT	3.00e+00
36	94	2.2	1 DDX8_CAEEL	PUTATIVE PRE-MRNA SPLI	1.56e+00
37	92	2.2	1 CO3_MOUSE	COMPLEMENT C3 PRECURSO	3.00e+00
38	91	2.2	1 CO3_CAVPO	COMPLEMENT C3 PRECURSO	4.14e+00
39	90	2.1	1 RL19_MYCPN	50S RIBOSOMAL PROTEIN	5.69e+00
40	90	2.1	1 AMEX_BOVIN	AMELOGENIN, CLASS I PR	5.69e+00
41	90	2.1	1 HA22_HUMAN	HLA CLASS II HISTOCOMP	5.69e+00
42	89	2.1	1 NIFS_ECOLI	NIFS PROTEIN HOMOLOG.	7.79e+00
43	89	2.1	1 RFBB_MYXXA	O-ANTIGEN EXPORT SYSTE	7.79e+00
44	89	2.1	1 RNRI_RAT	REGENERATING LIVER NUC	7.79e+00
45	90	2.1	1 POLR_KYMVJ	RNA REPLICASE POLYPROT	5.69e+00

ALIGNMENTS

RESULT 1
ID VA39_VACCC STANDARD: PRT; 403 AA.
AC P21062;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE PROTEIN A39.
GN A39R.
OS VACCINIA VIRUS (STRAIN COPENHAGEN).
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91021027.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PROLETTI E.;
RT "The complete DNA sequence of vaccinia virus."
RL VIROLOGY 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PROLETTI E.;
RL VIROLOGY 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M35027; G335517; -
DR PIR; E42521; E42521.
SQ SEQUENCE 403 AA; 45741 MW; FE0DB9AB CRC32;

Query Match 11.2%; Score 470; DB 1; Length 403;
Best Local Similarity 33.5%; Pred. No. 2.42e-86;
Matches 86; Conservative 59; Mismatches 95; Indels 17; Gaps 16;

Db	77	LVCOTNGNPKCWK	IGSDDPKHRGYPQNSKVTIISYN	ECVLSLDINISK	EG-I 132
Qy	109	LACGTNRHHPSCNVLNVTGVPVPLGEMRGYAFSPDENSEVSTIRKQYNGKI	168		
Db	133	KRWRRFDGPGCYDLYTADNVIPKDG	LRGAFVDKGDYD	KVYILFTDTICKSR	--I-VK 187
Qy	169	PRFRIRGE-S-ELYTSDTVNQNPQIFKATIVHQDQAYDDKIYFFREDNPKNPEAPLN	226		

QY 131 LGE-MRGYAPFSPDENSILVLFEGDEVYSTIRKQYNGK-IPRFRIRGESE-LYTSOTVM 187
Db 252 ES 253
QY 188 QN 189

RESULT 6
ID NRT1_RAT STANDARD; PRT; 275 AA.
AC P17982;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 (T-CELL MONO(ADP-
DE NAD(P)(+)-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL SURFACE PROTEIN
DE RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1)
GN AL2A OR RT6-A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEWIS A;
RX MEDLINE; 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell
RT alloantigen RT6.1";
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
RN [2]
RN MUTAGENESIS OF GLN-207.
RP MAEHAMA T., HOSHINO S.-I., KATADA T.;
RX MEDLINE; 96275529.
RA "Increase in ADP-ribosyltransferase activity of rat T lymphocyte
RT alloantigen RT6.1 by a single amino acid mutation.";
RL FEBS LETT. 388:189-191(1996).
CC -!- FUNCTION: HAS NAD+ GLYCOPOLYMERASE ACTIVITY AND EXTREMELY LOW ADP-
CC RIBOSYLTRANSFERASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
CC N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -!- CATALYTIC ACTIVITY: NAD(+) + H2O = NICOTINAMIDE + ADP-RIBOSE.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -!- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52082; G57168; -
CC EMBL; M31138; G206804; -
CC PIR; S08464; S08464.
CC PROSITE; PS01291; ART; 1.
CC PRAM; PF01129; ART; 1.
CC TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
KW T-CELL DIFFERENTIATION; GPI-ANCHOR.
FT SIGNAL 1 20 T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
FT CHAIN 21 245 HYDROPHOBIC, REMOVED DURING MATURATION
FT PROPEP 247 275 (BY SIMILARITY).
FT LIPID 246 246 GPI-ANCHOR (BY SIMILARITY).
FT ACT_SITE 209 209 BY SIMILARITY.
FT CARBOHYD 58 58 POTENTIAL.
FT MUTAGEN 207 207 Q-2E: INCREASED ADP-RIBOSYLTRANSFERASE
FT ACTIVITY.
SQ SEQUENCE 275 AA; 31388 MW; 523B1A84 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 2.88e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;
Db 194 LGVYIKFEFS-FYPDQEE-VLIPGVEYVQKVRTQGYNEIFLDSPKRKSNNCLYSSAGTR 251
QY 131 LGE-MRGYAPFSPDENSILVLFEGDEVYSTIRKQYNGK-IPRFRIRGESE-LYTSOTVM 187
Db 252 ES 253
QY 188 QN 189

RESULT 7
ID EF1A_TRIE STANDARD; PRT; 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEF1.
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OM9414 / RUTC-30;
RX MEDLINE; 94124021.
RA NAKARI T., ALATALO E., PENTTILA M.;
RT "Isolation of Trichoderma reesei genes highly expressed on glucose-
RT containing media: characterization of the tef1 gene encoding
RT translation elongation factor 1 alpha.";
RL GENE 136:313-318(1993).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-tRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z23012; G312887; -
CC PIR; S35772; S35772.
CC PROSITE; PS00301; EFATOR_GTP; 1.
CC PRAM; PF00009; GTP_EFTU; 1.
CC HSSP; P07157; IAIIP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 49830 MW; B9ABAB92 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 2.88e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;
Db 210 YKGWEKTKAGKTG-KTILEADSTDEPPKR-PTDKPLRLPLQDV 252
QY 478 YCGWDGRCISYSSERSVLSQINPAEPHKECPNPKAPLQKV 522

RESULT 8
ID VRP2_SALT STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

```

15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
65 KD VIRULENCE PROTEIN.
GN MKAA.
OS SALMONELLA TYPHIMURIUM.
OC PLASMID 96 KB VIRULENCE PEX102.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-TML R66;
RX MEDLINE; 90136009.
RA TAIRA S., RHEN M.;
RT "Identification and genetic analysis of mkaA -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth.";
RL MICROB. PATHOG. 7:165-173(1989).
RN [2]
RP SEQUENCE OF 1-10.
RX MEDLINE; 91244158.
RA TAIRA S., BAUMANN M., RIIKONEN P., SUKUPOLVI S., RHEN M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium.";
RL FEMS MICROBIOL. LETT. 61:319-323(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLA.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 215042; G47783; -.
DR PIR; A54540; A54540.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65440 MW; B69F25C3 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 2.88e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGLTAAARLSDPOAASHTAQM-LVESVTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RCENYITLLRRSEGLLACGTNARHPSCWNLVGTVPGLGMRGYAPFSPDENSILVLE 151
Db 220 GNEACGRDSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPFRIRGESELY 181

RESULT 9
ID VRP2_SALCH STANDARD; PRT; 591 AA.
AC P17450;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DE 65 KD VIRULENCE PROTEIN (PROTEIN M2, IN MBA REGION).
OS SALMONELLA CHOLERAESUIS (SALMONELLA ENTERICA).
OG PLASMID PKDSC50.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-RF-1;
RX MEDLINE; 90245675.
RA MATSUI H.;
RT "Nucleotide sequences of genes encoding 32 kDa and 70 kDa
RT polypeptides in mba region of the virulence plasmid, pKDS50, of
15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
65 KD VIRULENCE PROTEIN.
GN MKAA.
OS SALMONELLA TYPHIMURIUM.
OC PLASMID 96 KB VIRULENCE PEX102.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-TML R66;
RX MEDLINE; 90136009.
RA TAIRA S., RHEN M.;
RT "Identification and genetic analysis of mkaA -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth.";
RL MICROB. PATHOG. 7:165-173(1989).
RN [2]
RP SEQUENCE OF 1-10.
RX MEDLINE; 91244158.
RA TAIRA S., BAUMANN M., RIIKONEN P., SUKUPOLVI S., RHEN M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium.";
RL FEMS MICROBIOL. LETT. 61:319-323(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLA.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 215042; G47783; -.
DR PIR; A54540; A54540.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65440 MW; B69F25C3 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 2.88e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGLTAAARLSDPOAASHTAQM-LVESVTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RCENYITLLRRSEGLLACGTNARHPSCWNLVGTVPGLGMRGYAPFSPDENSILVLE 151
Db 220 GNEACGRDSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPFRIRGESELY 181

RESULT 9
ID VRP2_SALCH STANDARD; PRT; 591 AA.
AC P17450;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DE 65 KD VIRULENCE PROTEIN (PROTEIN M2, IN MBA REGION).
OS SALMONELLA CHOLERAESUIS (SALMONELLA ENTERICA).
OG PLASMID PKDSC50.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-RF-1;
RX MEDLINE; 90245675.
RA MATSUI H.;
RT "Nucleotide sequences of genes encoding 32 kDa and 70 kDa
RT polypeptides in mba region of the virulence plasmid, pKDS50, of

```

```

Salmonella choleraesuis.";
NUCLEIC ACIDS RES. 18:2181-2181(1990).
-!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
OF SALMONELLA.
-!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
PLASMIDS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X52035; G46898; -.
DR PIR; S09498; S09498.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65341 MW; BE4A439A CRC32;

Query Match 2.4%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 2.88e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGLTAAARLSDPOAASHTAQM-LVESVTPAGE-HIYYSYLAENGDNVDLN 219
QY 92 RCENYITLLRRSEGLLACGTNARHPSCWNLVGTVPGLGMRGYAPFSPDENSILVLE 151
Db 220 GNEACGRDSAMRYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPFRIRGESELY 181

RESULT 10
ID VRP2_SALEN STANDARD; PRT; 591 AA.
AC P55220;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN SPVB.
OS SALMONELLA ENTERITIDIS.
OG PLASMID PNL2001.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-AL1190;
RX MEDLINE; 94362897.
RA SUZUKI S., KOMASE K., MATSUI H., ABE A., KAWAHARA K., TAMURA Y.,
RA KIJIMA M., DANBARA H., NAKAMURA M., SATO S.;
RT "Virulence region of plasmid PNL2001 of Salmonella enteritidis.";
RL MICROBIOLOGY 140:1307-1318(1994).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLA.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; D14490; G517164; -.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 373 POLY-PRO.
SQ SEQUENCE 591 AA; 65350 MW; 9C5B2EE3 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 591;

```

```
Best Local Similarity 23.3%; Pred. No. 2.88e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHTAQW-LVEESVTPAGE-HIYISYLAENGDNVDLN 219
Qy 92 RCENYITLLERRSEGLACCTNARHPSCWNLVNGTVVPLGEMRGYAPSPDENSESLVFE 151
Db 220 GNEAGDRSAMRYSKV-QYGNATPAADLY 248
Qy 152 GDEVYSTIRKOEYNGKIPFRIRGESELY 181

RESULT 11
ID VRP2_SALDU STANDARD; PRT; 593 AA.
AC P24419;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN VSDC.
OS SALMONELLA DUBLIN.
OC PLASMID PSDL2.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91251759.
RA KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINEY D.;
RT "Molecular analysis of the virulence locus of the Salmonella dublin
plasmid PSDL2.";
RL MOL. MICROBIOL. 5:307-316(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L07305; G409249;
CC PROSITE: PS00154; ATPASE_E1_E2; 1.
CC PRAM; PF00122; E1-E2_ATPase; 1.
CC KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
ATP-BINDING.
CC FT DOMAIN 1 111 134 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 112 134 1 (POTENTIAL).
CC FT DOMAIN 135 136 1 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 137 155 2 (POTENTIAL).
CC FT DOMAIN 156 279 2 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 280 305 3 (POTENTIAL).
CC FT DOMAIN 306 320 4 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 321 350 4 (POTENTIAL).
CC FT TRANSMEM 351 711 5 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 712 734 5 (POTENTIAL).
CC FT DOMAIN 735 754 6 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 755 776 6 (POTENTIAL).
CC FT DOMAIN 777 817 7 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 818 843 7 (POTENTIAL).
CC FT DOMAIN 844 847 8 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 848 874 8 (POTENTIAL).
CC FT DOMAIN 875 916 8 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 916 916 8 ASP/GLU-RICH (ACIDIC).
CC FT MOD_RES 374 374 62 PHOSPHORYLATION (BY SIMILARITY).
CC FT BINDING 470 470 470 ATP (BY SIMILARITY).
CC FT SEQUENCE 916 AA; 98884 MW; 4A488A44 CRC32;
CC -----
Query Match 2.4%; Score 99; DB 1; Length 916;
Best Local Similarity 37.2%; Pred. No. 2.88e-01;
Matches 16; Conservative 8; Mismatches 16; Indels 3; Gaps 3;

Db 762 LAVGTWIT-ITTMLVGSENGIVONFORTHVPFLFLEISLTENW 803
Qy 244 LSVSKWNTFLKMLVCS-D-AATKNFNRLQDV-FLPDPSQW 284

RESULT 12
ID PMAL_AJECA STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
(DC-4 ALPHA CHAIN).
GN PMAL.
OS AJELLOMYCES CAPSULATA (HISTOPLASMA CAPSULATUM).
OS EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE: 84168117.
RA  AUFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPES D.,
RT  STROMINGER J.L.;
RT  "isotypic and allotypic variation of human class II
RL  histocompatibility antigen alpha-chain genes.";
RL  NATURE 308:327-333(1984).
RN  [2]
RP  SEQUENCE OF 40-255 FROM N.A.
RX  MEDLINE: 83065171.
RA  AUFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT  "cDNA clone for the heavy chain of the human B cell alloantigen DC1;
RT  strong sequence homology to the HLA-DR heavy chain.";
RL  PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: J00199; -; NOT_ANNOTATED_CDS.
CC  PIR: A02212; HLHUCD.
CC  PROSITE: PS00290; IG_MHC; 1.
CC  PFAM: PF00047; ig; 1.
CC  HSP: P01910; IIAK.
CC  MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT  SIGNAL 1 23
FT  CHAIN 24 255
FT  HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT  DQ(1) ALPHA CHAIN.
FT  DOMAIN 24 110
FT  EXTRACELLULAR ALPHA-1.
FT  DOMAIN 111 204
FT  EXTRACELLULAR ALPHA-2.
FT  DOMAIN 205 217
FT  CONNECTING PEPTIDE.
FT  TRANSMEM 218 240
FT  DOMAIN 241 255
FT  CYTOPLASMIC TAIL.
FT  DISULFID 133 189
FT  CARBOHYD 104 104
FT  POTENTIAL.
FT  CARBOHYD 144 144
FT  SEQUENCE 255 AA; 28105 MW; 5C46DD59 CRC32;
SQ  -----
Query Match 2.3%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1.12e+00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;

Db 25 DIVADSAQLG-VNLQSYGPGSQSHEDFGDEEFYVDLKERKTVQLPLFRFR 78
QY 122 NLVNGTVPLGEMRGYAPFSPDENSILVFEQDEV-YSTIRKQYNGKIPFRIR 175
-----
RESULT 14
ID VL2_HPVB5 STANDARD; PRT; 518 AA.
AC F26540;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MINOR CAPSID PROTEIN L2.
GN L2.
OS HUMAN PAPILLOMAVIRUS TYPE 5B.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91306467.
RA YABE Y., SAKAI A., HITSUMOTO T., KATO H., OGURA H.;
RT "A subtype of human papillomavirus 5 (HPV-5b) and its subgenomic
RT segment amplified in a carcinoma: nucleotide sequences and genomic
RT organizations.";
RL VIROLOGY 183:793-798(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90252; D1015003; -.
CC PIR: H40480; P2WL5.
CC PFAM: PF00513; late_protein_L2; 1.
CC COAT PROTEIN; LATE PROTEIN.
SQ SEQUENCE 518 AA; 56836 MW; D0ADC243 CRC32;
Query Match 2.3%; Score 95; DB 1; Length 518;
Best Local Similarity 27.1%; Pred. No. 1.12e+00;
Matches 16; Conservative 17; Mismatches 25; Indels 1; Gaps 1;

Db 219 GDITDIIELEEPSRYTFEIEPTPRRSSTPLPRNQSVGRRGFSLTNRRLVQOVQVD 277
QY 402 GTIHKVPEGEHSPAFNMEIQPFRAAAIQTMSLDAERRKLY-VSSQWEVSQVPLD 459
-----
RESULT 15
ID VL2_HPVO5 STANDARD; PRT; 518 AA.
AC P06318;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MINOR CAPSID PROTEIN L2.
GN L2.
OS HUMAN PAPILLOMAVIRUS TYPE 5.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87207670.
RA ZACHOW K.R., OSTROW R.S., FARAS A.J.;
RT "Nucleotide sequence and genome organization of human papillomavirus
RT type 5.";
RL VIROLOGY 158:251-254(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M17463; G484221; -.
CC PIR: B26277; P2WL5.
CC PFAM: PF00513; late_protein_L2; 1.
CC COAT PROTEIN; LATE PROTEIN.
SQ SEQUENCE 518 AA; 56704 MW; B7199004 CRC32;
Query Match 2.3%; Score 95; DB 1; Length 518;
Best Local Similarity 27.1%; Pred. No. 1.12e+00;
Matches 16; Conservative 17; Mismatches 25; Indels 1; Gaps 1;

Db 219 GDITDIIELEEPSRYTFEIEPTPRRSSTPLPRNQSVGRRGFSLTNRRLVQOVQVD 277
QY 402 GTIHKVPEGEHSPAFNMEIQPFRAAAIQTMSLDAERRKLY-VSSQWEVSQVPLD 459
-----
Search completed: Thu Jul 8 19:48:44 1999
Job time : 22 secs.

```

THIS PAGE BLANK (USPTO)

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	4201	100.0	666	4	075326	0.00e+00
2	2276	54.2	393	11	088371	0.00e+00
3	1850	44.0	653	14	064906	0.00e+00
4	737	17.5	748	11	062177	1.77e-15
5	725	17.3	751	11	062161	2.57e-15
6	716	17.0	749	4	013214	6.03e-14
7	716	17.0	750	4	033018	6.03e-14
8	710	16.9	751	4	099985	2.38e-14
9	705	16.8	751	13	042236	4.72e-14
10	688	16.4	753	4	013372	1.38e-14
11	682	16.2	754	11	088633	5.17e-14
12	681	16.2	772	13	030607	5.17e-14
13	682	16.2	785	11	088632	9.45e-14
14	681	16.2	785	4	013275	5.17e-14
15	679	16.2	785	4	Q15704	9.45e-14
16	669	15.9	772	11	062180	3.16e-13
17	665	15.8	772	11	062180	1.32e-13
18	656	15.6	771	4	Q63548	1.47e-13
19	618	14.7	782	11	Q45633	3.33e-13
20	614	14.6	775	11	062179	2.73e-12
						3.00e-12

Query Match	100.0%;	Score 4201;	DB 4;	Length 666;
Best Local Similarity	100.0%;	Pred. No. 0.00e+00;		
Matches	564;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Db	56	AVKKGHVGDQVDFGQTEPHTVLFPEPGSSVWVGGRGKVYLFDEPPEGKNASVURTWNIGS	115	
Qy	24	AVKKGHVGDQVDFGQTEPHTVLFPEPGSSVWVGGRGKVYLFDEPPEGKNASVURTWNIGS	83	
Db	116	TKGSCLDKRCENYITLLERRSEGLACCTNARHPCSWNLVNGTVVPLGEMRGYAPFSPD	175	
Qy	84	TKGSCLDKRCENYITLLERRSEGLACCTNARHPCSWNLVNGTVVPLGEMRGYAPFSPD	143	
Db	176	ENSLVLFEGDEVYSTTRKQYNGKIPFRFRIGESSELYTSDTVMQNPQIKATIVHQDQA	235	
Qy	144	ENSLVLFEGDEVYSTTRKQYNGKIPFRFRIGESSELYTSDTVMQNPQIKATIVHQDQA	203	
Db	236	YDDKIYFFREDNPKNPAPLNVSRVAQLCRDQGESSELSVSKWNTFLKAMLVCSDA	295	
Qy	204	YDDKIYFFREDNPKNPAPLNVSRVAQLCRDQGESSELSVSKWNTFLKAMLVCSDA	263	
Db	296	TNKNFNRLQDVFLLPDPGQWRDTRYGVFSNPWNYSAVCVYSLGDIQKVFRTSSLKGX	355	
Qy	264	TNKNFNRLQDVFLLPDPGQWRDTRYGVFSNPWNYSAVCVYSLGDIQKVFRTSSLKGX	323	

Db	356	SSLPNPRPGKCLPDQOPIPTETFOVADRHEVAQRVEPMGPKLTPLPHSKYHYQKVAVHR	415
Qy	324	SSLPNPRPGKCLPDQOPIPTETFOVADRHEVAQRVEPMGPKLTPLPHSKYHYQKVAVHR	383
Db	416	MQASHGETFHVLYLTTDRGTIHKVVEGEGEHSFAFNIMEIQPFRRAAAIQTMSLDAERR	475
Qy	384	MQASHGETFHVLYLTTDRGTIHKVVEGEGEHSFAFNIMEIQPFRRAAAIQTMSLDAERR	443
Db	476	KLYVSSQWEVSQVPLDCEVYGGCHCLMSRDPYCGWDGRCISIIYSSRSVLSQSNPA	535
Qy	444	KLYVSSQWEVSQVPLDCEVYGGCHCLMSRDPYCGWDGRCISIIYSSRSVLSQSNPA	503
Db	536	EPHKECNPKDPKAPLQKVLAPNSRYLLSCPMESRHATYSWRHKENVEQSCPEGHQSPN	595
Qy	504	EPHKECNPKDPKAPLQKVLAPNSRYLLSCPMESRHATYSWRHKENVEQSCPEGHQSPN	563
Db	596	CILFIENLTAQQYGHYFCEAQEGS	619
Qy	564	CILFIENLTAQQYGHYFCEAQEGS	587
RESULT 2			
ID	O88371	PRELIMINARY; PRT; 393 AA.	
AC	O88371		
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DE	SEMAPHORIN L (FRAGMENT).		
GN	SEMAL.		
OS	MUS MUSCULUS (HOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 98389619.		
RA	LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;		
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA		
RL	viruses.";		
RL	GENOMICS 51:340-350(1998).		
DR	ENBL; AF030699; G3523117; -.		
FT	NON_TER 393 393		
SQ	SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;		
Query Match 54.2%; Score 2276; DB 11; Length 393;			
Best Local Similarity 91.2%; Pred. No. 0.00e+00;			
Matches 311; Conservative 14; Mismatches 12; Indels 4; Gaps 2;			
Db	56	AVWKG---QDHVDSQPEPHTVLFHEPGSFVWVGGRGKYVHFNFPEGKNASVRTVNIGS	112
Qy	24	AVWKGHYGQDRVDFGQEPHTVLFHEPGSSVWVGGRGKYVLFDFPEGKNASVRTVNIGS	83
Db	113	TKGSCODKQDGGNYITLLERRGNGLLVCGTNARKPSCWNLVNDVSVNLSGEMKGYAPFSP	172
Qy	84	TKGCLDKRCCENITULLERRSEGLLACGNARHPSCWNLVNGTWV-PLGEMRGYAPFSP	142
Db	173	DENSLVLFEGDEVYSTIRKQYNGKIPFRIRIGESLYTSDTVMQNPQIKATIVHQDQ	232
Qy	143	DENSLVLFEGDEVYSTIRKQYNGKIPFRIRIGESLYTSDTVMQNPQIKATIVHQDQ	202
Db	233	AYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRDQGGESSLSVSKWNTFLKAMLVCSDA	292
Qy	203	AYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRDQGGESSLSVSKWNTFLKAMLVCSDA	262
Db	293	ATNRNFRNLQDVLLPDPSCQWRDTRYGVFSPNPNYSACVYSLGIDRIVFRFTSSLKGY	352
Qy	263	ATNKNFRNLQDVLLPDPSCQWRDTRYGVFSPNPNYSACVYSLGIDRIVFRFTSSLKGY	322
Db	353	HMGLSNRPGMCLPKKQPIPTETFOVADSHPEVAQRVEPMG	393
Qy	323	HSLPLNPRPGKCLPDQOPIPTETFOVADRHPEVAQRVEPMG	363

RESULT 3			
ID	Q64906	PRELIMINARY; PRT; 653 AA.	
AC	Q64906		
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)		
DT	01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)		
DE	SIMILAR TO GENBANK ACCESSION NUMBER L26081.		
OS	ALCELAAPHINE HERPESVIRUS 1.		
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;		
OC	GAMMAHERPESVIRINAE.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C500;		
RX	MEDLINE; 97201573.		
RA	ENSSER A., FLECKENSTEIN B.;		
RT	"Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";		
RL	J. GEN. VIROL. 76:1063-1067(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C500;		
RX	MEDLINE; 97404659.		
RA	ENSSER A., PELANZ R., FLECKENSTEIN B.;		
RT	"Primary structure of the alcelaphine herpesvirus 1 genome.";		
RL	J. VIROL. 71:6517-6525(1997).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C500;		
RL	SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.		
DR	EMBL; U18243; G1000717; -.		
DR	EMBL; AF005370; G2337970; -.		
SQ	SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;		
Query Match 44.0%; Score 1850; DB 14; Length 653;			
Best Local Similarity 47.5%; Pred. No. 0.00e+00;			
Matches 267; Conservative 95; Mismatches 188; Indels 12; Gaps 12;			
Db	86	GQHRF-FGQEPHTVLFHLSNDSVYVGGNNTIYLFDFAHSSNASTALNITSTHNLRL	144
Qy	31	GQDRVDFGQTEPHTVLFHEPGSSVWVGGRGKYVLFDFPEGKNASVRTVNIGTKGCLD	90
Db	145	SSTCENFIILLHNQDGLLACGTSNOKPSCW-LINLTTQFLGPKLGLAPFSSSNLVL	203
Qy	91	KRCENYITLLERRSEGLLACGNARHPSCWNLVNGTVVP-LGEMRGYAPFSSDENSLVL	149
Db	204	FDQNDYSTINLYKSLSGS-HKFRRIAGOVELYTSDTAMHRPOFVOATVHKNESYDDKI	262
Qy	150	FGDEVYSTIRK-QEYNGKIPFRIRIGESLYTSDTVMQNPQIKATIVHQDQATDDKI	208
Db	263	YFFFOENSHSDFKQPHPTVPRVGVQVCSQDGGESSLSVYKWTTFKARLACVDYDTGRIY	322
Qy	209	YFFREDNPDKNPEAPLNVSRVAQLCRDQGGESSLSVSKWNTFLKAMLVCSDAATKNF	268
Db	323	NEIQDIFTIHOAPENSWEETLIYGLFLSPNPFSAVCVTVTKVDIDHVEKTSKLNKHHKLP	392
Qy	269	NRLQDVLLPDPSCQWRDTRYGVFSPNPNYSACVYSLGIDRIVFRFTSSLKGYHSLPN	328
Db	383	PRPGCMCKNHOHVPTETFOVADRYPEVADPVYQKNNAMPEPIIQSKYIYTKLVYRVEYG	441
Qy	329	PRPGCLPQOQPIPTETFOVADRHPEVAQRVEPMGPKLTPLFHSHKYHYQKVAVHRQASH	388
Db	442	GVFWATIFYLTTIKGTIHYVRYEDNSNTALNILEINPFQKPAPIQNTILLDNLTNLKLYV	501
Qy	389	GETFHV-LVLYLTTDRGTIHKVVEGEGEHSFAFNIMEIQPFRRAAAIQTMSLDAERRKLYV	447
Db	502	NSWEVSEVPLDLCSVYGNDCFCFMSRDPCLCTWYNWTC-S-FK-QRVSVETGGGANRTL	558
Qy	448	SSQWEVSQVPLDLCEVYGGCHCLMSRDPYCGWDGRCISIIYSSRSVLSQSNPAEPH	506
Db	559	SEMCGDHYAPTIVVKHGVSTPLLSNLSVSLCPAVSNHADYFWTKDGFTEKRVHTKNDCI	618
Qy	507	KE-CPNPKDKAPLQKVLAPNSRYLLSCPMESRHATYSWRHKENVEQSCPEGHQSPNCI	565


```

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA YAMADA T., ENDO R., GOTOH M., HIROHASHI S.;
RT "Identification of semaphorin E as a non-MDR drug resistance gene of
RL PROC. NATL. ACAD. SCI. U.S.A. 94:14716-14718(1997).
DR EMBL: AB000220; D1033360; -.
SQ SEQUENCE 751 AA; 85207 MW; F89FB934 CRC32;

Query Match 16.9%; Score 710; DB 4; Length 751;
Best Local Similarity 33.6%; Pred. No. 2,28e-147;
Matches 151; Conservative 104; Mismatches 151; Indels 43; Gaps 30;

Db 167 SFNPVNTSVVINELEFSGMWI-DFMGTDAAIFRSLTKRNAVRTDQHNKSLWSEPMFVD 225
QY 139 PFPSPDENVLFEVGEDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNPFQIK 194
Db 226 AHVDPGDPNDKAYFFFEKELTDNNRSTKQIHSMIARICNDTGGRLSL-VNKNWTF 284
QY 195 ATIVHOD-QAYDDKIYFFREDNPKNEAPLVNSRVAQLCRDGGGESSLSVSKWNTFL 253
Db 285 KARLVCSYDDEGPTHEFDELEDFELL-ETDNP-RTTLVYGIFFTSVSSVFKGSVCVYHL 342
QY 254 KAMLVCS--DA-ATNKNENRLQDVELLPDPSQWRDTRVYGFSPWN-Y--SACVYSL 307
Db 343 SDIQTVFNGPFAHKEGPNHQLISYGRIPYRPGTCGGAFTPNMRTTKEFPDDVVTFIR 402
QY 308 GDIDKVFRTS-SLK-G-----YHSSLNPRPGKCLPDQ-QP-IPDET-F--QVAD-R 351
Db 403 NHPLMYNSIPIHKRLIVRI-GDYKYTKIAVDVNAADG-RYHVLFLGTDGTVQKV 460
QY 352 -HPEVAQRVEPM-G-PLKTPLEHSHYKQVAVHRMQASHGETTFVLYLTDDRGTHKV 408
Db 461 VLPTNNSVLELEFVFNKHAIPITTKISKKQQLYVSSNEGVSQVSLHRCRIYGT 520
QY 409 E-PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERKLYVSSQWESQVPLDCEVYGG 467
Db 521 CADCCCLARDPYCAWDGSCSRFYPTGKRKRSDQVHRGNPLTQCRGFNLKAYRNAAEIVQ 580
QY 468 HGCLMSRDPIYCGWDQGRICISYSE--RSVLQSNPAEPHKECP--NPKPKAPLQKVS 523
Db 581 YGVKNNTTFLECAPKSPQASIKWLLQDK 609
QY 524 LAP-NSRYLSCPMESRHATYSWR-HKEN 550

RESULT 9
ID O42236 PRELIMINARY; PRT: 751 AA.
AC O42236;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE COLLAPIN 3.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA FEINER L., KOPPEL A.M., KOBAYASHI H., RAPER J.A.;
RL NEURON 19:0-0(1997).
DR EMBL: AF022946; G2522204; -.
SQ SEQUENCE 751 AA; 85432 MW; 0B9BA2F7 CRC32;

Query Match 16.8%; Score 705; DB 13; Length 751;
Best Local Similarity 33.5%; Pred. No. 4,72e-146;
Matches 150; Conservative 107; Mismatches 150; Indels 41; Gaps 28;

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA XIANG R.H., HENSEL C.H., GARCIA D.K., CARLSON H.C., KOK K., DALY M.C.,
RA KERBACHER K., DEN BERG A., VELDHOUIS P., BUYS C.H., NAYLOR S.L.;
RT "Isolation of the human semaphorin iii/F gene (SEMA3F) at chromosome
RT 3p21, a region deleted in lung cancer."
RL GENOMICS 32:39-48(1996).
DR EMBL: U38276; G1061351; -.
DR PFW; PF00047; i9; 1.
SQ SEQUENCE 753 AA; 84941 MW; BECBEB0 CRC32;

Query Match 16.4%; Score 688; DB 4; Length 753;
Best Local Similarity 31.4%; Pred. No. 1,38e-141;
Matches 139; Conservative 121; Mismatches 140; Indels 42; Gaps 30;

Db 171 PYDPKLDTASALINEELVAGVYI-DFMGTDAAIFRTLGKQAMRTDQYNSRWLNDPSFI 229
QY 139 PFPSPDENVLFEVGEDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNPFQIK 194
Db 230 AELIPDSANDDKLYFFERSAE-APOSPAVYARIGICLNDGCHCL-VNKNWTF 287
QY 195 ATIVHODQAYDDKIYFFREDNPKNEAPLVNSRVAQLCRDGGGESSLSVSKWNTFL 254
Db 288 ARLVCSVPGEDGIEHFELODVY-QOTQDV-RNPVYAVFTSSGVSFRGSAVCVYNA 345
QY 255 AMLVCS---DAATNKNFRLQDVLDPDPSQWRDTRVYGFSPWN-WNY--SACVYSLG 308
```


RN SEQUENCE FROM N.A.
 RA ECKHARDT F., MEYERHANS A.;
 RT "Molecular cloning and expression pattern of a murine semaphorin
 homologous to H-sema IV";
 RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF080090; G337766; -;
 SQ SEQUENCE 785 AA; 88493 MW; 6FFCD8F5 CRC32;

 Query Match 16.2%; Score 682; DB 11; Length 785;
 Best Local Similarity 31.8%; Pred. No. 5.17e-140;
 Matches 141; Conservative 119; Mismatches 140; Indels 43; Gaps 31;

 Db 202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRTLGKOTAMRTDOYNSRWLNNDPSFIH 260
 QY 139 PFPSPDENSILVFEQDEYVSTIRKQYNGK-IPFRIRIGESELYTS--DTV-MONPOFIK 194
 Db 261 AELIPSAERNDDKLYFFFRERSAE-APONPAVYARIGRICLNDGHCCL-VNKNWSTFL 318
 QY 195 ATIVHQD-QAYDDKIYFFREDNPDKNPEAPLNVSVAQLCRDGGESSLSVSKWTF 253
 Db 319 KARLVCSVPGEDEIETHFDELQVYV--QOTQDI-RNPVIYAVFTSSGVSFRGSVAVCYVSM 376
 QY 254 KAMLVCS--DAATNKNFNRLQDVFLPDPSGOWRDRVYGVFSNP-WNY--SAVCVYSL 307
 Db 377 ADIRMVFNPFPAHKEGPNYQWMPFSGKMPYPRGTCGGTFTPSMKSTKDYDPDEVINMR 436
 QY 308 GDIDKVFRTS-SLK-G--YH-----SS-LPNRPCKLQDQ-QP-I-PTETF--QVAD--R 351
 Db 437 THPLMTQAVYPLQRRPLVVRT-GAPYRLTTVAVDQVDAADG-RYEVFLGCTDRGTQKVI 494
 QY 352 -HEVAQRVEPMG--PLKTPLFHSHKYQKVAVRHQASHGETHFVLYLTDRGTHKVV 408
 Db 495 VLPKDDQOEELMEEVEFEKAPYKVTMTISSKROOLYVASAVGVTHLSLHRCQAYGAA 554
 QY 409 E-PGEQEHSAFNIEMEIQPFRAAAIQTMSLDAERKLYVSSQWESVQVLDLCEVYGG 467
 Db 555 CADCCCLARDPYCWDGACSRYSYASSRRRRQDVHGNPIROCRGNSNANKNAVESVQ 614
 QY 468 CHCLMSRDPYCGWDGRCIS-IYSSER-SVLQSIINPAEPHKECP--NPKPDKAPLOKVS 523
 Db 615 YGVAGSAAFLECCPRSPQATVKW 637
 QY 524 LA-PNSRYILSCPMESRHATYSW 545

 RESULT 14
 ID Q13275 PRELIMINARY; PRT; 785 AA.
 AC Q13275; Q13274;
 DT 01-NOV-1996 (TREMREL. 01, CREATED)
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE SEMAPHORIN IV.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA NELSON J., BIEWALD T.;
 RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [1]
 RP SEQUENCE OF 394-436 FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE; 96210603.
 RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
 RA ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
 RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
 cancer deletion region and demonstrate distinct expression
 patterns";
 RT PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
 DR EMBL: AC000663; G1669373; -;
 DR EMBL: U32172; G995788; -;
 DR EMBL: U32171; G995786; -;

DR PFAM: PF00047; ig; 1;
 SQ SEQUENCE 785 AA; 88381 MW; CD175765 CRC32;

 Query Match 16.2%; Score 681; DB 4; Length 785;
 Best Local Similarity 31.4%; Pred. No. 9.45e-140;
 Matches 139; Conservative 122; Mismatches 139; Indels 43; Gaps 31;

 Db 202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRTLGKOTAMRTDOYNSRWLNNDPSFIH 260
 QY 139 PFPSPDENSILVFEQDEYVSTIRKQYNGK-IPFRIRIGESELYTS--DTV-MONPOFIK 194
 Db 261 AELIPSAERNDDKLYFFFRERSAE-APQSPAVYARIGRICLNDGHCCL-VNKNWSTFL 318
 QY 195 ATIVHQD-QAYDDKIYFFREDNPDKNPEAPLNVSVAQLCRDGGESSLSVSKWTF 253
 Db 319 KARLVCSVPGEDEIETHFDELQVYV--QOTQDI-RNPVIYAVFTSSGVSFRGSVAVCYVSM 376
 QY 254 KAMLVCS--DAATNKNFNRLQDVFLPDPSGOWRDRVYGVFSNP-WNY--SAVCVYSL 307
 Db 377 ADIRMVFNPFPAHKEGPNYQWMPFSGKMPYPRGTCGGTFTPSMKSTKDYDPDEVINMR 436
 QY 308 GDIDKVFRTS-SLK-G--YH-----SS-LPNRPCKLQDQ-QP-I-PTETF--QVAD--R 351
 Db 437 SHPLMTQAVYPLQRRPLVVRT-GAPYRLTTIAVDQVDAADG-RYEVFLGCTDRGTQKVI 494
 QY 352 -HEVAQRVEPMG--PLKTPLFHSHKYQKVAVRHQASHGETHFVLYLTDRGTHKVV 408
 Db 495 VLPKDDQOEELMEEVEFEKAPYKVTMTISSKROOLYVASAVGVTHLSLHRCQAYGAA 554
 QY 409 E-PGEQEHSAFNIEMEIQPFRAAAIQTMSLDAERKLYVSSQWESVQVLDLCEVYGG 467
 Db 555 CADCCCLARDPYCWDGACSRYSYASSRRRRQDVHGNPIROCRGNSNANKNAVESVQ 614
 QY 468 CHCLMSRDPYCGWDGRCIS-IYSSER-SVLQSIINPAEPHKECP--NPKPDKAPLOKVS 523
 Db 615 YGVAGSAAFLECCPRSPQATVKW 637
 QY 524 LA-PNSRYILSCPMESRHATYSW 545

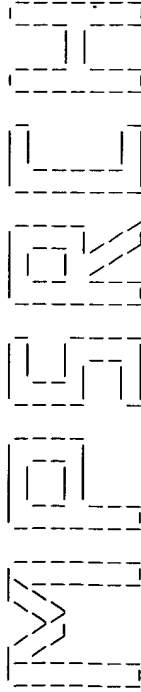
 RESULT 15
 ID Q15704 PRELIMINARY; PRT; 785 AA.
 AC Q15704;
 DT 01-NOV-1996 (TREMREL. 01, CREATED)
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE SEMAPHORIN.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRIN;
 RX MEDLINE; 96226360.
 RA ROCHE J., BOLDOG F., ROBINSON M., ROBINSON L., VARELLA-GARCIA L.,
 RA SWANTON M., WAGGONER B., FISHEL R., FRANKLIN W., GEMMILL R.,
 RA DRABKIN H.;
 RT "Distinct 3p21.3 deletions in lung cancer and identification of a new
 human semaphorin";
 RL ONCOGENE 12:1289-1297(1996).
 DR EMBL: U33920; G1000207; -;
 DR PFAM: PF00047; ig; 1;
 SQ SEQUENCE 785 AA; 88385 MW; 9971F14E CRC32;

 Query Match 16.2%; Score 679; DB 4; Length 785;
 Best Local Similarity 31.6%; Pred. No. 3.16e-139;
 Matches 140; Conservative 121; Mismatches 139; Indels 43; Gaps 31;

 Db 202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRTLGKOTAMRTDOYNSRWLNNDPSFIH 260
 QY 139 PFPSPDENSILVFEQDEYVSTIRKQYNGK-IPFRIRIGESELYTS--DTV-MONPOFIK 194
 Db 261 AELIPSAERNDDKLYFFFRERSAE-AQSPAVYARIGRICLNDGHCCL-VNKNWSTFL 318

QY 195 ATIVHOD-QAYDDKIYFFREDNPKNPEAPLNVSRVAQLCRGDOGESSLSVSKWNTFL 253
Db 319 KARLVCVPGEDGIEHFDLODFV-QOTODV-RNPVIYAVFTSSGVSFGRSACVYSM 376
QY 254 KAWLVCS---DAATNKNRLOQDFLLDPSPGQWRDTRVYGVFSPN-WNY--SACVYSL 307
Db 377 ADIRMYFNGSPFAHKEGPNYQWMPFGSKMYPYPRGTCPCGGTFPSMKSTKDYDDEVINEWR 436
QY 308 GDIDKVERTS-SLK-G--YH-----SS-LNPRPGKLPDQ-QP-I-PTETF--QVAD--R 351
Db 437 SHPLMYQVYPLQRRPLVVRT-GAPYRLTTIAVDQDAGDG-RYEVFLFGTDRGTQVKYI 494
QY 352 -HPEVAQRVEPMG--PLKTPLFHSHKYQKVAVHRMQASHGETFHVLYLTDRGTIHKVV 408
Db 495 VLPKDDQEMEELMLEVEVEFKDPAPYKMTIISKROOLYVASAVGVTHLSLHRCQAYGAA 554
QY 409 E-PGEGEHSFAFNIMEIQPFRAAAIQTMSLDAERRKLYSSQWEYSQVPLDLCEYGGG 467
Db 555 CADCCCLARDPYCAWDQACSRYTASSKRSRRDVRHGNPIROCRGFNSNANKNAVESYQ 614
QY 468 CHCCLMSRDPYCGWDGRCIS-IYSSER-SVLOSINPAEPHKECP--NPKPKAPLOKYS 523
Db 615 YGVAGSAAFLECOPRSPQATVKW 637
QY 524 LA-PNSRYVYLSCPMESRHATYSW 545

Search completed: Thu Jul 8 19:49:55 1999
Job time : 53 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:55:07 1999; Maspar time 27.36 Seconds
Tabular output not generated. 432.983 Million cell updates/sec

Title: >US-09-041-236-2

Description: (12-568) from US09041236.pep (42 of 45)

Perfect Score: 4151
Sequence: 1 AQGHLRSGPRIFAWKRGVH.....ENVEQCEPGHQSPNCILFI 557Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.083; Variance 147.463; scale 0.245

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	656	15.8	771	13	Human semaphorin III	2.30e-52
2	615	14.8	477	13	Human collapsin.	3.17e-48
3	608	14.6	775	33	Human semaphorin.	1.61e-47
4	606	14.6	861	22	Mouse Cbl100 antigen.	2.56e-47
5	599	14.4	861	32	Human semaphorin.	1.30e-46
6	538	13.0	862	22	Human Cbl100 antigen.	1.71e-40
7	518	12.5	776	32	Rat semaphorin W.	1.70e-38
8	463	11.2	441	13	Vaccinia virus semaphorin	4.97e-33
9	459	11.1	730	13	Grasshopper semaphorin	1.24e-32
10	428	10.3	587	32	Human semaphorin W.	1.42e-29
11	390	9.4	974	33	Tribolium secreted protei	7.52e-26
12	378	9.1	712	13	Tribolium semaphorin	1.11e-24
13	367	8.8	650	13	Drosophila semaphorin	1.31e-23
14	362	8.7	724	13	Drosophila semaphorin	3.99e-23
15	352	8.5	930	32	Human semaphorin Y.	3.72e-22
16	325	7.8	888	25	Human semaphorin Z.	1.48e-19

17	316	7.6	887	25	W19856	Rat semaphorin Z.	1.08e-18
18	296	7.1	929	32	W57259	Rat semaphorin Y.	8.71e-17
19	157	3.8	122	13	R71385	Variola major virus s	4.00e-04
20	99	2.4	591	4	R23006	Protein transcribed f	1.82e+01
21	98	2.4	832	33	W61092	Taq DNA polymerase I	2.16e+01
22	101	2.4	2329	25	W25038	Partial BRCA2 cancer	1.30e+01
23	96	2.3	832	33	W61090	Taq DNA polymerase I	3.02e+01
24	91	2.2	478	1	R04881	Recombinant elastase.	6.92e+01
25	91	2.2	598	9	R48631	Sequence of nuclear r	6.92e+01
26	90	2.2	638	5	R03924	E. coli HSP (dnaK).	8.15e+01
27	92	2.2	832	33	W61091	Taq DNA polymerase I	5.87e+01
28	92	2.2	854	30	W56309	Clas II S-receptor ki	5.87e+01
29	93	2.2	855	6	R23815	S receptor kinase pro	5.87e+01
30	93	2.2	4572	30	W52845	A. mediterranei rifam	4.98e+01
31	89	2.1	143	25	W25052	BRCA2 cancer suscepti	9.59e+01
32	89	2.1	534	25	W25031	Partial BRCA2 cancer	9.59e+01
33	89	2.1	680	4	R23143	Mutant thermostable D	9.59e+01
34	89	2.1	757	4	R23141	Mutant thermostable D	9.59e+01
35	89	2.1	787	4	R23142	Mutant thermostable D	9.59e+01
36	89	2.1	832	13	R76691	DNA-polymerase REM-T3	9.59e+01
37	89	2.1	832	12	R64272	T. aquaticus DNA-poly	9.59e+01
38	89	2.1	832	1	P90556	Purified native therm	9.59e+01
39	89	2.1	832	33	W61089	Taq DNA polymerase I	9.59e+01
40	89	2.1	832	13	R76690	Taq DNA-polymerase RE	9.59e+01
41	89	2.1	833	27	W24211	Cleavage DN nuclease.	9.59e+01
42	89	2.1	833	27	W24212	Cleavage DN nuclease.	9.59e+01
43	89	2.1	833	27	W24213	Cleavage DV nuclease.	9.59e+01
44	89	2.1	3418	26	W19211	Human breast cancer s	9.59e+01
45	89	2.1	3418	26	W23287	Human breast and ovar	9.59e+01

ALIGNMENTS

RESULT 1

ID R71380 standard; Protein; 771 AA.

AC R71380; 771

DT 21-NOV-1995 (first entry)

DE Human semaphorin III protein.

KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;

KW variola major virus; smallpox; semaphorin receptor binding activity;

KW modulation; nerve cell growth; immune response; viral pathogenesis;

KW neurological disease; neuro-regeneration; oncological infection.

OS Homo sapiens.

PN WO9507706-A.

PD 23-MAR-1995.

PF 13-SEP-1994; U10151.

PR 13-SEP-1993; US-121713.

PA (REGC) UNIV CALIFORNIA.

PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;

PI O'Connor T;

DR WPI; 95-131177/17.

DR N-PSDB; Q87442.

PT New class of semaphorin peptide(s) and polypeptide(s) - are

PT potent modulators of nerve cell growth and regeneration

PS Example 2; Page 60-63; 101pp; English.

CC The sequence of the human semaphorin III protein.

CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,

CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II

CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)

CC virus semaphorin IV (Q87447) genes were used to generate a series of

CC peptides (R70370-R70418), which retain semaphorin receptor binding

CC activity. The semaphorin derived or semaphorin receptor binding

CC are potent modulators of nerve cell growth, immune responsiveness and

CC viral pathogenesis. They can be used in diagnosis and treatment of

CC neurological disease and neuro-regeneration, immune modulation and

CC diagnosis and treatment of viral and oncological infection and diseases.

SQ Sequence 771 AA;

Query Match

Best Local Similarity 15.8%; Score 656; DB 13; Length 771;

Matches 143; Conservativity 31.2%; Pred. No. 2.30e-52;

Matches 143; Mismatches 115; Indels 45; Gaps 35;

Db 166 rgkspdykbltasllidgelysgta-adfmgdrdfairtighhhpirtqhsrwnldp 224

Db 462 kvitlyngqetwmeevileelqifkdpapilsmieisskrqglyiqsasavagvrfhcdm 521
 QY 406 KVEPEGEHSAFN-IMP-IOFFRAAAIQMSDAERKKLYVSSQWEVSQVPLDLCEV 463
 Db 522 ygsacacclardpycawdgiscrypt 550
 QY 464 YGGGCHGCLMSRDPYCGWDOGRCSISYSS 492

RESULT 4

ID W17658 standard; Protein: 861 AA.
 AC W17658;
 DT 24-JUL-1997 (first entry)
 DE Mouse CD100 antigen.
 KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
 OS Mus sp.
 FH key
 FT peptide
 FT protein
 FT protein
 FT domain
 FT domain
 FT domain
 FT domain
 FT domain
 FT domain
 FT domain
 FT modified_site
 FT W09717368-A1.
 FT 15-MAY-1997.
 FT 09-NOV-1996; U18645.
 FT (DAND) DANA FARBEN CANCER INST.
 FT Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
 FT WPI: 97-280982/25.
 FT N-PSDB: T60666.
 FT Nucleic acid molecule encoding CD100 antigen - which stimulates
 PT leukocyte response, e.g. B cell aggregation, differentiation,
 PT survival and T cell proliferation
 PS Example 8: Page 86-89; 135pp; English.
 CC Mouse CD100 antigen (W17657) is a novel leukocyte semaphorin-like
 CC protein that stimulates a leukocyte response, including B cell
 CC aggregation, B cell differentiation, B cell survival and/or T cell
 CC proliferation. Its amino acid sequence was deduced from a cDNA
 CC clone (T60666) isolated from murine T cells. Human CD100 antigen
 CC (W17657) has also been identified. CD100 polypeptides and fusion
 CC proteins, nucleic acids, and host cells expressing CD100 can be
 CC utilised in diagnostic and therapeutic methods involving modulation
 CC of B and T cell responses, neuron axonal growth and immune cell-
 CC nerve cell interaction.
 SQ Sequence 861 AA;

Query Match 14.6%; Score 606; DB 22; Length 861;
 Best Local Similarity 32.4%; Pred. No. 2.56e-47;
 Matches 143; Conservative 106; Mismatches 145; Indels 47; Gaps 31;

Db 107 eclnyrvlqlsstslvycgtfnafqctdhlnltsfklgsedgkrcpfdphsyt 166
 QY 93 DCENYITLLER-RSEGLACGTNARHPSC--WNLVNGTVPVLGEM-RGYAPFSPDENSIV 148
 Db 167 vmvggelysgts-yfnlgsepiisrnsnshsplrteyaipwnepsfvfadvikspdppe 225
 QY 149 LFEQDEVYSTTKQYNGKIPFRRIERGESELYTSDTV--MONPOFIKATIVHQ--D--Q 202
 Db 226 geddkvyfftevsveyefvfkmlprvarvckgqggglrtlq-kkwtstflkarlicskp 284

QY 203 AYDDKIYFFREDNPKNPEAPLVNSRVAQLCRGQGGESSLSVSKWNTFLKAMLVCSDA 262
 Db 285 dsglvfnllqdvflrap-gl-kepvfyavfptqinnvglsavcaytlatveavfsrky 342
 QY 263 ATKNFNRLQDVFLLPDPGQWRDTRVYGFSPNPNY---SAVCVYSLGDIDKVF-R--- 315
 Db 343 mqsatvegshktwryngpvtprpgacidsearaanytsslnlpdkltqlfvkdhplmd 402
 QY 316 --TSSL-----K-G-YHSSLNPNRPKGKCL-PDQQPIP-TETFQVADRHP-E-V-AQR-VE- 360
 Db 403 svtpidnrpklikkdvnvtqivvdrtdqldgtfydvmmfistdrgalghkaviltkev--v 460
 QY 361 PMGPLKT-P-LFHSKYHYQKVAVHRMQASHGETFVILVLTTRDGTIRHKVVEGEGEHFA 418
 Db 461 --ieetqlfrdseplvtlllsskgrkfvyagsngvvgaplafcekhgs-ceedvlar 517
 QY 419 FNIMEIQPFERRAAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDLCEVYGGGCHGCLMSRD 476
 Db 518 pycawspaikacvtlhqeeas 538
 QY 477 PYCGWDOG-R-CISYSSERS 495

RESULT 5

ID W58540 standard; Protein: 861 AA.
 AC W58540;
 DT 02-SEP-1998 (first entry)
 DE Human semaphorin.
 KW Human semaphorin; diagnosis; nervous disease; immune disease;
 KW nerve extension inhibiting activity.
 OS Homo sapiens.
 FN J10155490-A.
 PD 16-JUN-1998.
 PF 27-NOV-1996; 332900.
 PR 27-NOV-1996; JP-332900.
 PA (SUMO) SUMITOMO SEIYAKU KK.
 DR WPI: 98-391044/34.
 DR N-PSDB: V31121.
 PT New human semaphorin gene - useful in the diagnosis of nervous system
 PT and immune disorders
 PS Claim 1; Page 10-12; 15pp; Japanese.
 CC The present sequence represents human semaphorin (translated from the
 CC Japanese specification as semaforin). Semaphorin has nerve extension
 CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
 CC treatment and researches on nervous diseases and immune diseases.
 SQ Sequence 861 AA;

Query Match 14.4%; Score 599; DB 32; Length 861;
 Best Local Similarity 32.4%; Pred. No. 1.30e-46;
 Matches 143; Conservative 105; Mismatches 146; Indels 47; Gaps 31;

Db 107 eclnyrvlqlsstslvycgtfnafqctdhlnltsfklgsedgkrcpfdphsyt 166
 QY 93 DCENYITLLER-RSEGLACGTNARHPSC--WNLVNGTVPVLGEM-RGYAPFSPDENSIV 148
 Db 167 vmvggelysgts-yfnlgsepiisrnsnshsplrteyaipwnepsfvfadvikspdppe 225
 QY 149 LFEQDEVYSTTKQYNGKIPFRRIERGESELYTSDTV--MONPOFIKATIVHQ--D--Q 202
 Db 226 geddkvyfftevsveyefvfkmlprvarvckgqggglrtlq-kkwtstflkarlicskp 284
 QY 203 AYDDKIYFFREDNPKNPEAPLVNSRVAQLCRGQGGESSLSVSKWNTFLKAMLVCSDA 262
 Db 285 dsglvfnllqdvflrap-gl-kepvfyavfptqinnvglsavcaytlatveavfsrky 342
 QY 263 ATKNFNRLQDVFLLPDPGQWRDTRVYGFSPNPNY---SAVCVYSLGDIDKVF-R--- 315
 Db 343 mqsatvegshktwryngpvtprpgacidsearaanytsslnlpdkltqlfvkdhplmd 402
 QY 316 --TSSL-----K-G-YHSSLNPNRPKGKCL-PDQQPIP-TETFQVADRHP-E-V-AQR-VE- 360
 Db 403 svtpidnrpklikkdvnvtqivvdrtdqldgtfydvmmfistdrgalghkaviltkev--v 460


```
Oy 194 KATIHQDQAYDDKIYFFREDNPD-KNPEAPLNVSRVAQLCGDGGESSLSVSKWNTF 252
Db 262 lksrlncsvpgdyfpyfneiqstdliegnvggq-vekllygvtftpvnsgsgscafs 320
Oy 253 LKAMLVCSDAATNK-NFNRLQDVF-LLDPP-SQWRDTRVYGFSPNPN-Y--SACVYS 306
Db 321 mksilesfdgfkqetmnsnvlavpslkvppeprgqcvndrtlpdvsvnfvkshtlmd 380
Oy 307 L-G--D-IDKVFRT--SSLKG-VHS--SL--PNPRPGKCLPDQOPIPTETFOVADRPEVA 356
Db 381 eavpafft-rplllislsqyrtfklavdqqvtpdgkaydvlfigtddgkvikalnasaf 439
Oy 357 QRVEPMGLKTPLFH-S-KYHYQKAV-HRMOASGETFHVLYLTDRGTTHKVVPEG-- 411
Db 440 dssdsvdsvieelqvlppgvpknyvrmgdgddsklvvssddelaiklhrsgskit 499
Oy 412 EGEHSEAFNIME-IQFRRAAAIQTM---SLDAERKLYVSSQWEVSQVPLDLCVEYG-G 466
Db 500 ncrecvslqdpvcavdnvelketavgspdw 530
Oy 467 GCHGCLMSRDPYCGWDQ-G-RCISIVSSERS 495

RESULT 10
ID W51314 standard; Protein: 587 AA.
AC W51314;
DT 08-SEP-1998 (first entry)
DE Human semaphorin W.
KW Human; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PI Kikuchi K, Kimura T;
PI W09815628-A1.
PD 16-APR-1998
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287836.
PR (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
PI WPI; 98-261015/23.
DR N-PSDB; V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 10.3%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 1.42e-29;
Matches 108; Conservative 89; Mismatches 149; Indels 32; Gaps 22;

Db 57 dedgdelyffttetsrafdsyerikvprvarvcagdlgqrktlq-qrwttfkladllcp 115
Oy 201 DOAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCGDGGESSLSVSKWNTFLKAMLVCS 260
Db 116 gpehgrassvldgavlrlpelga-gtptfygiffssqwegatisavcafrpqdirtvlnpg 174
Oy 261 DAATKNFNRLQDVFLLPDPSCQWRDTRVYGFSPNPN-Y--SACVYSLGDIDRVRTS 317
Db 175 frelkhdcnrglpvvdndvpqpgccitnnmklrhfsslspldrvitfirhplmdrp 234
Oy 318 --SLK-GYHSSLP--N---PRPGKCLPDQOPIPTETFOVADRPEV-V-AQR-VE-P 361
Db 235 vfpagdhplllvtdtaylrvvahrvtlsqkeydvlylgtedghlhravrigag-ls-vl 292
Oy 362 MGPLKT-PLFHSK-RHYQKAVHRMOASGETFHVLYLTDRGTTHKVVPEGEQHSFAF 419
Db 293 edlal--fpepqpvenmkl-yhsw-llvgstrtevtqnttcnrglqs-csecilaqdpvc 347
```

```
Oy 420 NIMETQPFRAAAIQTMSLDAERRKLYVSSQWEVSQVPLDLCVEYGGCHGLMSRDPYC 479
Db 348 awsfrldecvahagehglvcdiesadvsslcpkqogerpvfvfepvataahvlpccps 407
Oy 480 GWD-Q-GRCISIVSSERSVLOSINPAEPHEKCPNPKPKAPLQKVLAPNSRYLYLSCPME 537
Db 408 sawascvvhqpsgv 421
Oy 538 SRHATYSWRHKENV 551

RESULT 11
ID W64221 standard; Protein: 974 AA.
AC W64221;
DT 06-OCT-1998 (first entry)
DE Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
PI WPI; 98-362774/31.
DR N-PSDB; V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 174; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 9.4%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 7.52e-26;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechfnfklvllknddalfvcgtnafnpscrnykmdtlepfdfgdfsgmarcpydakan 172
Oy 91 KRDCENYIT-LLERRSEGLACGTNARHPSCNWLVNGTVPLG-EMRGYA--PESPENS 146
Db 173 valfadgklysatvt-dflaidaviysl-gesp--tlrtvkhskwlkpeyfvqadv 228
Oy 147 LVLFEGDEVYITRKQYENG-KIPFRFRIRGESELYTSDTYMQNPQFIKATIVHQDQAYD 205
Db 229 dylyfffreiaevyntgkvfvrvagvckndmgsgvleqkqtsfklarincsvpgds 288
Oy 206 DKIIYFFREDNPDKNPEAPLNVSRVAQLCRDQGGESSLSVSKWNTFLKAMLVCSDAATN 265
Db 289 hfyfnlqvadvlring--rdv-vlatfstpynsigpavcaydmldiasvftgrfkeq 345
Oy 266 K-NFNRLQDVFLLPDPSCQWRDTRVYGFSPNPN-Y--SACVYSLGIDIKVFE--RTSSL 319
Db 346 kspdstwtvpdervpvrprccagsssleryatsnefpddtlfnfikhplmdaevpsif 405
Oy 320 KGYHSS---LPNPRPGKCLPD--QOPIPTETFOVADRPEVA-Q--RVEP-M-G-P-L- 365
```


OY 471 CLMSRDPYCGWQGR-CISIYSSE-RSVLQSIINPAE-PHKECPN 511

Search completed: Thu Jul 8 19:55:43 1999
Job time : 36 secs.

THIS PAGE BLANK (USPTO)

WQSEFH
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 19:54:07 1999; Maspar time 24.09 Seconds
Tabular output not generated. 926.287 Million cell updates/sec

Title: >US-09-041-236-2
Description: (12-568) from US09041236.pap (42 of 45)
Perfect Score: 4151
Sequence: 1 AQGHLRSGPRIFAVWKGHV.....ENVQSCPEGHQSPNCILFI 557

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 48.785; Variance 86.556; scale 0.564

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	1830	44.1	553	2 T03102	semaphorin homolog A3	0.00e+00
2	737	17.8	748	2 I48744	semaphorin A - mouse	6.05e-134
3	725	17.5	751	2 I48748	semaphorin E - mouse	3.40e-131
4	716	17.2	749	2 G01856	semaphorin V - human	3.90e-129
5	688	16.6	753	2 G02173	semaphorin III family	9.69e-123
6	681	16.4	772	2 A49069	collapsin - chicken	3.83e-121
7	669	16.1	772	2 I48747	semaphorin D - mouse	2.07e-118
8	662	15.9	666	2 I58169	semaphorin III - mouse	8.10e-117
9	656	15.8	771	2 D49423	semaphorin III precursor	1.87e-115
10	618	14.9	782	2 I48746	semaphorin C - mouse	7.76e-107
11	522	12.6	834	2 S65498	M-sema F protein precursor	2.83e-85
12	474	11.4	760	2 I48745	semaphorin B - mouse	1.25e-74
13	470	11.3	403	2 E42521	A39R protein - vaccinia	9.51e-74
14	463	11.2	441	2 D29921	hypothetical protein	3.31e-72
15	461	11.1	730	2 JH0798	fasciclin IV precursor	9.11e-72
16	448	10.8	1074	2 JCS528	semaphorin F precursor	6.51e-69
17	389	9.4	711	2 A49423	semaphorin I precursor	4.35e-56
18	367	8.8	656	2 B49423	semaphorin I - fruit	2.28e-51
19	362	8.7	724	2 C49423	semaphorin II precursor	2.66e-50
20	293	7.1	295	2 JQ1775	Sall9R protein - vacc	7.86e-36
21	196	4.7	1884	2 JC4975	plexin 2 precursor -	1.13e-16
22	170	4.1	1872	2 JC4976	plexin 3 precursor -	6.34e-12
23	171	4.1	1905	2 I51553	Plexin - African claw	4.21e-12

24 168 4.0 1894 2 JC4980 plexin 1 precursor - 1.44e-11
25 157 3.8 122 2 JQ1845 l4R protein - variola 1.21e-09
26 157 3.8 122 2 H36852 A43R protein - variol 2.67e-09
27 155 3.7 142 2 JQ1776 Sall9R protein - vacc 1.27e-03
28 120 2.9 775 2 E70320 polyribonucleotide nu 1.90e-01
29 105 2.5 227 2 S09922 hypothetical protein 1.90e-01
30 105 2.5 235 1 QOBEC9 HXL4 protein - human 1.90e-01
31 102 2.5 275 2 A34866 T-cell surface protei 4.88e-01
32 103 2.5 406 2 B69064 conserved hypothetical 3.57e-01
33 99 2.4 275 2 S08464 T-cell alloantigen R 1.23e+00
34 99 2.4 460 2 S35772 translation elongatio 1.23e+00
35 99 2.4 563 2 S78224 virulence-associated 1.23e+00
36 99 2.4 591 2 S26565 virulence-associated 1.23e+00
37 99 2.4 591 2 S09498 virulence-associated 1.23e+00
38 99 2.4 591 2 S22664 virulence-associated 1.23e+00
39 99 2.4 593 2 S15215 virulence-associated 1.23e+00
40 99 2.4 597 2 A46050 thyroid/steroid recep 1.23e+00
41 98 2.4 641 2 S32017 flagellum-associated 1.66e+00
42 98 2.4 970 2 C57282 ankyrin-related prote 1.66e+00
43 98 2.4 979 2 B57282 ankyrin-related prote 1.66e+00
44 99 2.4 1375 2 JC5148 hepatocyte growth fac 1.23e+00
45 98 2.4 1786 2 A57282 ankyrin-related prote 1.66e+00

ALIGNMENTS

RESULT 1

ENTRY T03102 #type complete
TITLE semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM #formal_name alcelaphine herpesvirus 1
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999

ACCESSIONS T03102
REFERENCE 214840
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal J. Virol. (1997) 71:6517-6525
#title Primary structure of the alcelaphine herpesvirus 1 genome.
#accession T03102

##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-653 #label ENS
##cross-references EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY #length 653 #molecular_weight 73645 #checksum 5501

Query Match 44.1%; Score 1830; DB 2; Length 653;
Best Local Similarity 47.0%; Pred. No. 0.00e+00;
Matches 264; Conservative 98; Mismatches 188; Indels 12; Gaps 12;

Db 67 AKSRFDKPRILVNLTDGFGQHRF-FGQPEHTVLFHSLNSSDVYVGGNNTIYLFDFAH 125
Qy 12 AQGHLRSGPRIFAVWKGHVGDVDFGQTEPHTVLFHEPGSSVWVGGRKYLDFDFEG 71
Db 126 SNAATALINITSTHNRSLTSCTENFILLNQTDLGACGTSNOKPSCW-LINNLTQF 184
Qy 72 KNASVTVNIGSTRGSLDKDCENYITLLERSGSLACGTSNARHPSCWNLVNGTVVP- 130
Db 185 LGPKLGLAPFSSPNLVLFQNDTYSTINLYKSUGS-HKFRRIAGQVELYTSDTAMHR 243
Qy 131 LGEMRGYAPFSPDENSILVLFEGDEVYSTRK-QEYNGKIPRFRIRGESELYSDTVMQN 189
Db 244 PQFVQATAVHNESYDDKIYFFQENSHSDKQFPHTVPRYGVQVCSDDGGESSLSVYKW 303
Qy 190 PQFIKATIVHQDQYDDKIYFFREDNPDKNPEAPLNSRVAQLCRGGOGESSLSVSKW 249
Db 304 TTFKLARLACVDYDGRYINELQDIFIWQAPENSWEETLIYGLFLSPWNSFACVFTVKD 363
Qy 250 NTFKLAMLVCSDAATKNKFNRLQDVLFPDPSGGQRDTRVYGVFSNPWNYSAVCYISGD 309
Db 364 IDHVEKTSKLNHYHKLTPRPGQCMKHQHVPTTFQVADRYPEVADPVYQKNNAMEPI 423
Qy 310 IDKVFRTSSLKGYHSSLNPRPGKCLPQDQPIPTTFQVADRHPEVAORVPMGPKLTPL 369
Db 424 IQSKIIYTKLLVRYVEYG-GVFWATIFYLTIKGTIHIYRYEDSNSTALNILEINPQ 482

```
Qy 370 FHSKYHYQVAVHRMQASHGETFH-VLTLTDRGTIHKVPEPGEQHSFAENIMEIQFER 428
      :|:|:| | | :|:|:| | | :|:|:| | | | :|:|:| | | |
Db 483 KPAPTIONLLDNTNLKLYVNSEWESEVPLDLCSYVGNDCSCFMSRDPCLCTWYNTTC-S 541
      :|:|:| | | | | | :|:|:| | | | | | :|:|:| | | | | |
Qy 429 RAAAIQTMSLDAERKLYVSSQWSEVPLDLCEVYGGCHGCLMSRDPYCGWQGRGIS 488
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 542 -FK-QRVSVETGGPANRTLSEMGDGHAPTIVVKVQVSIPLLSNYLSCPAVSNHADFWT 599
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 489 IYSSERSVLOSINPAEPH-KE-CNPXPDKAPLOKVSLAPNSRYLSCPMSRHTATYSWR 546
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 600 KDGFEKTRCHVYTHKNOCILLI 621
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 547 HKNVQSCPCPGHSPNCILFI 568
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |

RESULT 2
ENTRY semaphorin A - mouse
TITLE semaphorin A - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48744
REFERENCE #type complete
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48744
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues 1-748 #label RES
#cross-references EMBL:X85990; NID:g854323; PID:g854324
GENETICS semA
#gene semA
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 748 #molecular-weight 82894 #checksum 9017

Query Match 17.8%; Score 737; DB 2; Length 748;
Best Local Similarity 33.7%; Pred. No. 6.09e-134;
Matches 168; Conservative 112; Mismatches 166; Indels 52; Gaps 36;

Db 112 ECMNFVRLHAYNTHLLACRTGAFHPTCALWRWATAGTGHASTGPEKLEDGKGTYPDP 171
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 93 DCENYITLLERRSEG-LLACCTNARHSC-----NNLVNQT-V-V-P-LGEMRGYAFESP 142
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 172 RHRPSPVLVBELYSGV-TADLMGRDFTIFRSLGONPSLTERPHDSRWLNPNPKVFWFI 230
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 143 DENSLVLFEGDEVYSTIRKQYNGK-IPRFRIRGESELYT-S-DTV-MONPOFIKAT-I 197
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 231 PESENPDKKIYFFRESAVAAAPAMGMSVSRVQICRNDLGGQSL-VNKNWTFILKAR 289
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 198 VHQAQVDDKIYFFREDNPDKNPE-APLNVSRYAQLCRGQGGESSLSVSKNFTFLKAM 256
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 290 LVCSVPVGEVGHFDQLOQVFLLS-SR-DROTPLLYAVFSTSSGVFGSACVYVSMNDVR 347
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 257 LVCS-DAATNKNFNRLQDVFLLPDPSCQWRDTRVYGVFSNPNW-Y-SAVCVYSLGDID 311
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 348 RAFGLGPLHKEGPHQWVSYGVRVYPRPGMCPKSTGTGTSSTKDFDDVYQFGRNHPLM 407
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 312 KVFRTS-SLK-G-----YHSSLNPNRPKGKCLPDQ-QPIP-TETF--QVAD--R-HPEV 355
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 408 YNPVLPKGG-R-PLFLQVAGYTTQTAADRAVAADGH-YDVLFIGTDVGTVLKVISVPK 464
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 356 AQRVPEPMGLKTLPLF-H--SKYHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVWE-P- 410
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 465 GRPNSEGLLEELQVPEDSAAITSMOISSKQOOLYASRAVAQIALHRCALTALGRACAE 524
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 411 GEQHSFAFNIMEIQPFRAAAIQTMSLDAERKLYVSSQWSEVPLDLCEVYGGCHG 470
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 525 CCLARDPYCAWDGSGACTRFQPTAKRRFRQDRIKNDPSTLCSGSSHSVLEKKVLGVES 584
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
```

```
Qy 471 CLMSRDPYCGWQDQGRGISIYSSERSVL--QSINPAEPHKECPNPKDKAPLOKVSLAPNS 528
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 585 GSAFLECEPESLOAHVQM 602
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 529 -RYLSCPMSRHTATYSW 545
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |

RESULT 3
ENTRY semaphorin E - mouse
TITLE semaphorin E - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48748
REFERENCE #type complete
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48748
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues 1-751 #label RES
#cross-references EMBL:X85994; NID:g854331; PID:g854332
GENETICS semE
#gene semE
CLASSIFICATION #superfamily semaphorin
SUMMARY #length 751 #molecular-weight 85259 #checksum 8961

Query Match 17.5%; Score 725; DB 2; Length 751;
Best Local Similarity 33.0%; Pred. No. 3.40e-131;
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;

Db 167 SFNPVNTVSMINEELFSQWYI-DFMGTDAAIFRSUTKRMQLRDTDOHNSKWLSEPFVD 225
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 139 PFPSPDENSLVLFEGDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MONPOFIK 194
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 226 AHVIPDGTDPNDKAVYFFKERTLDNNRSTKQIHSMTARICPNDTGGQSL-VNKNWTFEL 284
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 195 ATIVHQD-QAYDKIYFFREDNPDKNPEAPLNVSRYAQLCRGQGGESSLSVSKNFTFL 253
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 285 KARLVCSVTDEDEPETHFELEVDVFL-ETDNP-RITLVYGIETTSVSVFKGSACVYVHL 342
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 254 KAMLVCS-DA-ATNKNFNRLQDVFLLPDPGQWRDTRVYGVFSNPNW-Y-SAVCVYSL 307
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 343 SDIOTVFNGBFAKEGPNHOLISYQGRIPYPRPGTCPPGGAFTPNMRTTKDFDDVVTFR 402
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 308 GDIDKVFRTS-SLK-G-----YHSSLNPNRPKGKCLPDQ-QP-I-PTETF--QVADRHP 353
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 403 NHPLMYSISPIHRRPLIVRIGTDYKYTKTAVDRVNAADG-RYHVLFLGTDRTGVQKVYV 461
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 354 EVAQRVPEPMGLK-TPLF-H--SKYHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVWE 409
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 462 LPNASSAGSGLILEELVFNKHVPIITMEISSKKOOLYVSSNEGVSQVSLHRCHIYGTAC 521
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 410 -PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERKLYVSSQWSEVPLDLCEVYGGC 468
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 522 ADCCCLARDPYCAWDGSHCSRFYPTGKRKRSDQVDRHGNPLTQCRGFNLKAYRNAAEIVQY 581
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 469 HGCLMSRDPYCGWQDQGRGISIYSS--RSVLQINPAEPHKECP--NPKPKAPLOKVSL 524
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Db 582 GVNNSTFLECAPKSPQASIKWLLQDK 609
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |
Qy 525 A-PNSRYILSCPMSRHTATYSWR-HKEN 550
      :|:|:| | | :|:|:| | | :|:|:| | | :|:|:| | |

RESULT 4
ENTRY semaphorin V - human
TITLE semaphorin V - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
04-Sep-1998
```

Matches	139;	Conservative	121;	Mismatches	140;	Indels	42;	Gaps	30
Db	171	PYPDKLDTASALINEELYAGYII--DFMGTDAAIFRTLGKOTAMRTDQYNSRWLNDRPSFIH	229						
QY	139	PFPDENSLVLFEGDEYVIRKQYNGK-IPFRRTIRGESELYTS--DIV-MQNPOFIK	194						
Db	230	AELIPDSAENDKLYIFFERSAE-AQSPAVYARIGRICLNDDGGHCL-VNKMSTFLK	287						
QY	195	ATIVHQDAQYDDKIYFFERDNPDKNEAPLNVSRVAQLCRGQDQGESSLVSKWNTFLK	254						
Db	288	ARLVCSVPGEDGTHETHEDELQDFV-QQTDV-RNPVIYAVFTSSGVSFRGSACVYVMSA	345						
QY	255	AMLVCS---DAATNKNFNRLQDFLLPDPGQWRDTRVGVFSNP-WNY--SAYCVYSLG	308						
Db	346	DIRVFNFGPAHREGPNYQWMPFSGKMPYPRPGFCPGTTPSKMKSIXDYDPVEINEMRS	405						
QY	309	DIDKVFRTS-SLK-G-YH---SS-LPNRPGKCLPDQ-QP-I-PTETF--QVAD--R-	351						
Db	406	HPLMYQAVYQLQRRLPVVRT-GAPIRLTTIAVDQVDSADG-RYEVLFLGTRDGTQVKVIV	463						
QY	352	HPEVAQRVEPMG-PLKTLFHSKYHYKVAVHRMQASHGETFHVLYLTTRDGTIHKVVE	409						
Db	464	LPKDDQMEELMLEEVEVFQDPAPVKWTMTSSKRQQLYVASAVGVTHLSLHRQAYGAAC	523						
QY	410	-PGEQSHFAFNIEMIQPFERRAAAIQMTSDAERKKLYVSSQWEVSQVPLDLCEVYGGC	468						
Db	524	ADCLLARDPYCANDGOACSRYTASSKRSRRODVRHGNPIROCGRFNSNANKNAVESVOY	583						
QY	469	HGCLMSRDPYCGMDQGRGIS-IYSSEY-SVLQSLNPAEPHKECP--NPKDPKAPLQKSL	524						
Db	584	GVAGSRAFLFCQPRSPQATVKW	605						
QY	525	A-PNSRYILSCPMESRHATYSW	545						
RESULT	6	A49069	#type complete						
ENTRY		collapsin - chicken							
TITLE		#formal_name Gallus gallus							
ORGANISM		07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change							
DATE		04-Sep-1998							
ACCESSIONS		A49069							
REFERENCE		A49069							
#authors		Luo, Y.; Raible, D.; Raper, J.A.							
#journal		Cell (1993) 75:217-227							
#title		Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones.							
#accession		A49069							
#status		preliminary: not compared with conceptual translation							
#molecule_type		mRNA							
#residues		1-772	#label LUO						
#cross-references		GB:002528; NID:g410078; PID:g410079							
CLASSIFICATION		#superfamily semaphorin							
SUMMARY		#length 772 #molecular-weight 88867 #checksum 9712							
Query Match		16.4%; Score 681; DB 2; Length 772;							
Best Local Similarity		31.5%; Pred. No. 3.8e-121;							
Matches		146; Conservative 119; Mismatches 150; Indels 48; Gaps 38							
Db	166	RGKSPYDPKLLTASLLVDGELYSGTA-ADPMGRDFAIFRTLGHHPHPIRTEQHOSRWLNDRP	224						
QY	135	RGVAFSPDENSLVLFEGDEVISTIRKQYNGK-IPFRIRGESELYTS-DIV-MQNPN	190						
Db	225	RFTSAHLIPSDNPEDDKIYFFFRFNNAIDEGHTGKATHARIGQICKNDFGHRSL-VNKKW	283						
QY	191	QFIKATIVHO-DQAYDDKIYFFFRDNPNKPEAPLNVSRVAQLCRGQDQGESSLVSKW	249						
Db	284	TTFLKARLICSVPGNPIDTHFEDELQDFVLM-NSKDP-KNPIYGVFTTSSNIFKGSVAC	341						
QY	250	NTEFLKMLVCS--DA-ATNKNFNRLQDFVLLPDPGQWRDTRVGVFSNPWN-Y--SAYC	303						
Db	342	MYSMTDVRRVFLGPIYAHHRDGNPQWVYQGRVYPRPGTSPKSTFGGFDSTKDLDPDEVIT	401						

```
QY 304 VYSLGIDKVERTS-SLK-G--YH--S--S-LPNPRGKCLPDQ-QPI-PTETF--QV-A 349
Db 402 FARSHAMYNPVFPINS-REPIMIKTDVDYQFTQIVVDRVDAEDGO-YDVMFTGTDIGTVL 459
QY 350 -DR-HPEVAQRVPMGLKPTPLFHSK--YHYQKVAVRHQASHGETFHVLYLTDRGTI 405
Db 460 KVSIPKETWHEELLEEMTVFREPTVISAMKISTKQOOLYIGSATGVSQSLPLHRCDV 519
QY 406 KVE-PGEQHSFAFIME-IOPFRAAAIQTMSLDAERRKLYVSSQWESVQVPLDCEV 463
Db 520 YGKACAECLARDPYCANDGSSCSRFPPTAKRTRRQDTRNGDPLTHCSDLOLQHDNPSQ 579
QY 464 YGGGCHGLMSRDPYCGWQGRGICISYSE-RSVL-QSINPAEPHKECPN-PKPDKAP-- 518
Db 580 TLEEKIYGVENSSTLECSPKSQRAIVY-WQFQKONDDHKVE 621
QY 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVQSCE 556

RESULT 7
ENTRY I48747 #type complete
TITLE semaphorin D - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
ACCESSIONS I48747
REFERENCE I48744
#authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
#title family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48747
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-772 #label RES
#cross-references EMBL:X85993; NID:g854329; PID:g854330
GENETICS
#gene semD
#superfamily semaphorin
CLASSIFICATION #length 772 #molecular_weight 88710 #checksum 1776
SUMMARY
Query Match 16.1%; Score 669; DB 2; Length 772;
Best Local Similarity 30.2%; Pred. No. 2,07e-118;
Matches 170; Conservative 143; Mismatches 182; Indels 67; Gaps 51;

Db 58 HTFLDE-ERSRLYGAKHIFSFNLVNIKDFOKIWPVSY-TRRDECKWAGKDIKECA 115
QY 43 HTVLFHEPGSSSVWVGGRGVLYLDFPEGKN-AS-VRTVNIQSTKGC--LDK--RDCE 95
Db 116 NETKVLKAYNOTHLXACGTGAHPICITYIEVGHHPEDNIFKLODSHFENGKSPYDPKL 175
QY 96 NYITLLERSEG--LLACGNTNARHPSCWNL-V--N--GTVVPL--G--EM-RGVAPESPDE 144
Db 176 LTASLLIDGELYSGTA-ADFMGRDFAIFRTLGDHHPHPIRTEQHDNRWLNDRPRTSARLIPE 234
QY 145 NSLVFEGDEVYSTIRKQYNGK-IPFRIRGESELYTS--DTV-MQNPQIKATIVHQ 200
Db 235 SDNPEDKVIYFFPRENAIGEGHSGKATHARIGQICKNDFGGHRS-L-VNKWTFLLKARLIC 293
QY 201 -DOAYDDKIYFFREDNPDKNPEAPLVNSRVAQLCRGDQGGESSLSVSKWNTFLKAMLYC 259
Db 294 SVPGPNGLDTHDELDQVFLM-NSKDP-KNPIYVGVFTTSSNIFKGSACMYSMDSVRV 351
QY 260 S--DA-ATNKNFNRLQDVFLLDPSPQGWDRTRVYGVFSNPWN-Y--SACVYVSLGDIDKV 313
Db 352 FLGPYAHROGPNYQWVPYQGRVPYPRPGTCSKTFGFGDSTKDLDPDVIIFGRSHPAMXN 411
QY 314 FRTS-SLK-G--YH--S--S-LPNPRGKCLPDQ-QPI-PTETF--QV-A-DR-HPEVAQ 357
Db 412 PVFPINN-RPIMIKTDVNYQFTQIVVDRVDAEDGO-YDVMFTGTDVGTVLKVVSPKPTW 469
QY 358 RVEPMGLKPTPLFHS--KYHYQKVAVRHQASHGETFHVLYLTDRGTIHKVVE-PGEQ 414
```

```
Db 470 HDLEEVLLEEMTVFREPTTISAMELSTKQOOLYIGSTAGVAOLPLHRCDDIYKCAECCL 529
QY 415 HSFAFNIME-IOPFRAAAIQTMSLDAERRKLYVSSQWESVQVPLDCEVYGGGCHGLM 473
Db 530 ARDPYCANDGSSCSRFPPTAKRTRRQDTRNGDPLTHCSDLEDHNDHNGHGPSLEERIIYGV 589
QY 474 SRDPYCGWQGRGICISYSE-RSVL-QSINPAEPHKECPN-PKPDK--AP-LQ-KVSLA- 525
Db 590 ENSSTFLECSPKSQRALVY-WQ 610
QY 526 PNSRYLSC-PMESRHATYSWR 546

RESULT 8
ENTRY I58169 #type fragment
TITLE semaphorin III - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
ACCESSIONS I58169
REFERENCE I58169
#authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;
#journal Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#title Semaphorin III can function as a selective chemorepellent to
#title pattern sensory projections in the spinal cord.
#cross-references MUID:95267432
#accession I58169
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-666 #label RES
#cross-references GB:L40484; NID:g703189; PID:g703190
GENETICS
#gene SemalIII
#superfamily semaphorin
CLASSIFICATION #length 666 #checksum 9654
SUMMARY
Query Match 15.9%; Score 662; DB 2; Length 666;
Best Local Similarity 30.7%; Pred. No. 8.10e-117;
Matches 141; Conservative 122; Mismatches 149; Indels 48; Gaps 37;

Db 60 RGKSPYDPKLLTASLLIDGELYSGTA-ANFMGRDFAIFRTLGHHPHPIRTEQHDNRWLNDRP 118
QY 135 RGYAPFSDENSLVLFEGDEVYSTIRKQYNGK-IPFRIRGESELYTS--DTV-MQNP 190
Db 119 RPSAHLIPESDNPEDDKVYFFPRENAIGEGHSGKATHARIGQICKNDFGGHRS-L-VNKW 177
QY 191 QPIKATIVHQ-DOAYDDKIYFFREDNPDKNPEAPLVNSRVAQLCRGDQGGESSLSVSKW 249
Db 178 TTFELKARLICSVPGPNGLDTHDELDQVFLM-NSKDP-KNPIYVGVFTTSSNIFKGSACV 235
QY 250 NTFELKAMLYCS--DA-ATNKNFNRLQDVFLLDPSPQGWDRTRVYGVFSNPWN-Y--SACV 303
Db 236 MYSMDSVRVLLGPYAHROGPNYQWVPYQGRVPYPRPGTCSKTFGFGDSTKDLDPDVIIT 295
QY 304 VYSLGIDIKV----F--RTSS-LK--GYHSSLNPNRPGKCLPDQ-QPI-PTETF--QV-A 349
Db 296 FARSHAMYNPVFPINN-RPIMIKTDVNYQFTQIVVDRVDAEDGO-YDVMFTGTDVGTVL 353
QY 350 -DR-HPEVAQRVPMGLKPTPLFHS--KYHYQKVAVRHQASHGETFHVLYLTDRGTI 405
Db 354 KVSVPKETHWDLLEEVLEEMTVFREPTTISAMELSTKQOOLYIGSTAGVAOLPLHRCDI 413
QY 406 KVE-PGQEHSAFIME-IOPFRAAAIQTMSLDAERRKLYVSSQWESVQVPLDCEV 463
Db 414 YGKACAECLARDPYCANDGSSCSRFPPTAKRTRRQDTRNGDPLTHCSDLOLQHDNHP 473
QY 464 YGGGCHGLMSRDPYCGWQGRGICISYSE-RSVL-QSINPAEPHKECPN-PKPDK--AP 518
Db 474 SLEERIYGVENSSTFLECSPKSQRALVY-WQFORNEDR 512
QY 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQ 553
```


Search completed: Thu Jul 8 19:54:49 1999
Job time : 42 secs.

THIS PAGE BLANK (USPTO)

(TM)

Result No.	Query %			ID	Description	Pred. No.
	Score	Match	Length			
1	470	11.3	403	1	VA39_VACCC	3,78e-86
2	463	11.2	441	1	VA39_VACCV	2,37e-84
3	163	3.9	1871	1	SEX_HUMAN	8,46e-13
4	105	2.5	227	1	US08_HCMVA	1,50e-02
5	102	2.5	275	1	NRT2_RAT	3,02e-01
6	99	2.4	275	1	NRT1_RAT	1,02e-01
7	99	2.4	460	1	EF1A_TRTRE	2,93e-01
8	99	2.4	591	1	VRP2_SALTY	2,93e-01
9	99	2.4	591	1	VRP2_SALCH	2,93e-01
10	99	2.4	591	1	VRP2_SALEN	2,93e-01
11	99	2.4	593	1	VRP2_SALDU	2,93e-01
12	99	2.4	916	1	PNAL_AJECA	2,93e-01
13	98	2.4	1132	1	DNBI_HSV60	4,13e-01
14	95	2.3	255	1	RS21_HUMAN	1,14e+00
15	94	2.3	329	1	RESR_ECOLI	1,14e+00
16	95	2.3	518	1	VL2_HPV5B	1,58e+00
17	95	2.3	518	1	VL2_HPV05	1,14e+00
18	95	2.3	683	1	AMOH_ARTGO	1,14e+00
19	94	2.3	947	1	PNW2_YEAST	1,58e+00
20	94	2.3	959	1	N100_YEAST	1,58e+00
21	94	2.3	1069	1	ENTK_MOUSE	1,58e+00
22	94	2.3	1115	1	IREI_YEAST	1,58e+00
23	94	2.3	1200	1	DDX8_CAEEL	1,58e+00

Db	280	TILYVFEDSPYSKASALCTYSMTIKOSFSTKLEGYTKOLPSPASGICILPACKVVPHTTF	339
Qy	287	TRYGVFSFWNTSYACVYVSLGIDIKVFTSSLKYGHSSLPNRPCKLUPDOOPIETF	346
Db	340	EVIEKYNVLDDIILKPLS	356
Qy	347	QVADRHPEVAORVEPMG	363

01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
TRANSMEMBRANE PROTEIN SEX PRECURSOR.
GN
OS
OC
OC
OC
OC
GN
SEX.
HOMO SAPIENS (HUMAN).
EUKARYOTA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
SEQUENCE FROM N.A.
TISSUE=FETAL BRAIN, SKELETAL MUSCLE, AND EMBRYO;
MEDLINE: 96149362.
RA MAESTRINI E., LONGATTI P., CREMONA O., GULISANO M.,
RA BIONE S., TAMANINI F., NEEL B.G., TONIOLO D., COMOGGIO P.M.;
RT "A family of transmembrane proteins with homology to the
RT MET-hepatocyte growth factor receptor."
RL PROC. NATL. ACAD. SCI. U.S.A. 93:674-678 (1996).
CC -1- FUNCTION: PUTATIVE RECEPTOR INVOLVED IN THE DEVELOPMENT OF NEURAL
CC AND EPITHELIAL TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -1- TISSUE SPECIFICITY: PREDOMINANT IN NEURAL TISSUE DURING
CC DEVELOPMENT (BY SIMILARITY); WIDELY EXPRESSED IN ADULT TISSUE.
CC -1- SIMILARITY: WITH THE EXTRACELLULAR DOMAIN OF THE MET/ROS/HGF
CC RECEPTORS AND THE CYTOPLASMIC DOMAIN OF SEP, NOV AND OCT.

CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL outstat-
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commo-

EMBL; X87852; E183847; -
 DR MIM; 300022; -
 DR RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
 KW SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1871 TRANSMEMBRANE PROTEIN SEX.
 CC

FT	CARBOHYD	738	738	POTENTIAL.
FT	CARBOHYD	746	746	POTENTIAL.
FT	CARBOHYD	1009	1009	POTENTIAL.
FT	CARBOHYD	1036	1036	POTENTIAL.

FT CARBOHYD	1073	1073	POTENTIAL.
FT CARBOHYD	1115	1115	POTENTIAL.
FT CARBOHYD	1162	1162	POTENTIAL.
SQ SEQUENCE	1871 AA	207661 MW;	9A11046A CRC32;

Query Match	3.9%;	Score 163;	DB 1;	Length 1871;
Best Local Similarity	29.5%;	Pred. No. 8.46e-13;		
Matches	26;	Conservative	29;	Mismatches 29;
				Indels 4;
				Gaps 3;

Db 427 VVFIGTRSGSLKKVRVDGQDAHL-YETVPYVD--GSPILRDLLFSPDRHIYLLSEKQV 483
|:::| |::| |::| |::| |::| |::| |::| |::| |::|
Qy 394 VLYLTDRGTIHKVVEGGEHSAFNIMETQPFRRAAAIQNSLDAERKKLYVSSQWEV 453

QY	131	LGE-MRCYAPSPDENSESLVLFEGDEVYSTRKQYNGK-IPRERRIRGESE-LYTSDTVW	18
DB	252	ES 253	
QY	188	QN 189	
RESULT	6		
ID	NRT1_RAT	STANDARD;	PRT: 275 AA.
AC	PI7982;		
DT	01-NOV-1990	(REL. 16, CREATED)	
DT	01-NOV-1990	(REL. 16, LAST SEQUENCE UPDATE)	
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)	
DT	T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1	PRECURSOR (EC 2.4.2.31) (T-CELL	
DE	NAD(P)(-)-L-ARGININE ADP-RIBOSYLTRANSFERASE 1	(T-CELL MONO(ADP-	
DE	RIBOSYL)TRANSFERASE 1	(ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN	
DE	RT6.1).		
GN	ART2A OR RT6-A.		
OS	RATTUS NORVEGICUS (RAT).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;		
OC	RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.		
RC	[1]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=Lewis A;		
RX	MEDLINE; 90192088.		
RX	HAAG F., KOCH F., THIELE H.-G.;		
RA	"Nucleotide and deduced amino acid sequence of the rat T-cell		
RT	alloantigen RT6.1.";		
RL	NUCLEIC ACIDS RES. 18:1047-1047(1990).		
CC	[2]		
CC	MUTAGENESIS OF GLN-207.		
CC	MEDLINE; 96275529.		
RA	MAEHANA T., HOSHINO S.-I., KATADA T.;		
RT	"Increase in ADP-ribosyltransferase activity of rat T lymphocyte		
RT	alloantigen RT6.1 by a single amino acid mutation.";		
RT	PBS LEFT. 388:189-191(1996).		
CC	-1- FUNCTION: HAS NAD+ GLYCOHYDROLASE ACTIVITY AND EXTREMELY LOW ADP-		
CC	RIBOSYLTRANSFERASE ACTIVITY.		
CC	-1- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +		
CC	N2-(ADP-D-RIBOSYL)-L-ARGININE.		
CC	-1- CATALYTIC ACTIVITY: NAD(+) + H(2)O = NICOTINAMIDE + ADP-RIBOSE.		
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.		
CC	-1- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.		
CC	-1- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).		
CC	-1- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE		
CC	FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	tion between the Swiss Institute of Bioinformatics and the EMBL Outstat-		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for comm-		
CC	entities requires a license agreement (See http://www.isb-sib.ch/anno/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X52082; G57168;		
DR	EMBL; M31136; G206804;		
DR	PIR; S08464; S08464.		
DR	PROSITE; PS01291; ART; 1.		
DR	PFAM; PF01129; ART; 1.		
KW	TRANSFERASE; GLYCOSYLTRANSFERASE; GPI-ANCHOR.		
KW	T-CELL DIFFERENTIATION; GPI-ANCHOR.		
FT	SIGNAL	1 20	
FT	CHAIN	21 246	
FT	PROPEP	247 275	
FT	LIPID	246 246	
FT	ACT SITE	209 209	
FT	CARBOHYD	58 58	
FT	MUTAGEN	207 207	
FT	ACTIVITY.		
QO	SEQUENCE	275 AA; 31388 MW; 523B1A84 CRC32;	

```

Query Match      2.4% Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred.No. 2.93e-01;
Matches         19; Conservative          17; Mismatches 21; Indels    5; Gaps   5;

Db 194 LGVIKESF-FYPDQEE-VLIPIGYEYVQKVTGYNIEFLDSPKRKSNVCILYSAGTR 251
|| : :: | ||:: ||: ||| ||| : ||| : : ||: ||: |||
Oy 131 LGE-MRGYAPSPDENSLVFEGDEVISTIRKOYEYNGK-IPFRRIERGESE-LYTSDTVM 187
|| : :: | ||:: ||: ||| ||| : ||| : : ||: ||: |||

Db 252 ES 253
|| : :: | ||:: ||: ||| ||| : ||| : : ||: ||: |||

Oy 188 QN 189

RESULT 7
ID EFIA_TIRE STANDARD; PRT; 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEFI.
OC TRICHODERMA REESEI (HYPOCREA JECORINA),
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
CN HYPOCREALES; HYPOCREACEAE; HYPOCREA.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-QM9414 / RUTC-30;
RX MEDLINE: 94124021.
RA NAKARI T., ALATALO E., PENTTILA M.;
RT "Isolation of Trichoderma reesei genes highly expressed on glucose-
RT containing media: characterization of the tef1 gene encoding
RT translation elongation factor 1 alpha.";
RL GENE 136:313-318(1993).
CC -! FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-tRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -! SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -! SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on i
CC use by non-profit institutions as long as its content is in no way m
CC modified and this statement is not removed. Usage by and for commerci
CC entities requires a license agreement. (See http://www.isb-sib.ch/announc
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; Z23012; G312887; -.
CC DR PIR; S35772; S35772.
CC DR PROSITE; PS00301; EFATOR_GTP; 1.
CC DR PFAM; PF00009; GTP_EFTU; 1.
CC DR HSPD; P07157; IAI_P.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT PP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 49830 MW; B9ABAB2 CRC32;

Query Match      2.4% Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred.No. 2.93e-01;
Matches         15; Conservative          13; Mismatches 15; Indels    2; Gaps   2;

Db 210 YKGWEKETRACKTG-KTLLEADTSIEPPKR-PTDKPLRLPLODV 252
| ||| : :: : ||: ||: ||| : ||| : : ||: ||: |||
Oy 478 YCGWDQGRCSISYSSERSVLQSINPAEPHKCPNPDKAPLOKV 522
| ||| : :: : ||: ||: ||| : ||| : : ||: ||: |||

RESULT 8
ID VFP2_SALT_Y STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

```

[illegible]


```

RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE: 95266321.
RA  GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA  MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;
RT  "The DNA sequence of human herpesvirus-6: structure, coding content,
RT  and genome evolution.";
RL  VIROLOGY 209:29-51(1995).
CC  -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC  REPLICATION.
CC  -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC  -!- SIMILARITY: TO OTHER HERPESVIRUSES' DNA-BINDING PROTEIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: X92436; G1044871; -.
DR  EMBL: X83413; G854020; -.
DR  PFAM: PF00747; Viral_DNA_bp: 1.
KW  DNA-BINDING; DNA REPLICATION; ZINC-FINGER; NUCLEAR PROTEIN.
FT  2N-FING 459 475 C4-TYPE.
SQ  SEQUENCE 1132 AA; 127762 MW; DB699A02 CRC32;
-----
Query Match 2.4%; Score 98; DB 1; Length 1132;
Best Local Similarity 28.2%; Pred. No. 4.13e-01;
Matches 24; Conservative 18; Mismatches 37; Indels 6; Gaps 6;
-----
Db 435 WNLNRNRYVYNAGNAHTEI-YNHLVNCANLCEFCDCGC-C-QSCIGTAMRVGTRLPALP 491
QY 297 WYSAVCYVSLGSDIDKVFRTSSLGKHYHSLNPRPGKCLPDQOPIPTETFOVADRHEVA 356
-----
Db 492 KNVK-KEPLVMSMF-SRY-YAEVDI 513
QY 357 QRVEPMPKLPFLFHSKYHYQVAV 381
-----
RESULT 14
ID HA21_HUMAN STANDARD; PRT; 255 AA.
AC P01908.
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84168117.
RA AUFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPEL D.,
RA STROMINGER J.L.;
RT "Isotypic and allotypic variation of human class II
RT histocompatibility antigen alpha-chain genes.";
RL NATURE 308:327-333(1984).
RN [2]
RP SEQUENCE OF 40-255 FROM N.A.
RX MEDLINE: 83065171.
RA AUFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT "cDNA clone for the heavy chain of the human B cell alloantigen DC1:
RT strong sequence homology to the HLA-DR heavy chain.";
RL PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93278300.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural homology between tbs repressor and ribose binding protein
RT implies functional similarity.";
RL PRONIN SCI. 1:843-849(1992).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RSDACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING
```

```

CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: J00199; -. NOT_ANNOTATED_CDS.
DR  PIR: A02212; HLHUDC.
DR  PROSITE: PS00290; IG_MHC; 1.
DR  PFAM: PF00047; Ig_1.
DR  PFAM: PF00993; MHC_II_alpha; 1.
DR  HSP: P01910; IIAK.
KW  MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT  SIGNAL 1 23
FT  CHAIN 24 255
FT  DOMAIN 24 110
FT  DOMAIN 111 204
FT  DOMAIN 205 217
FT  TRANSMEM 218 240
FT  DOMAIN 241 255
FT  DISULFID 133 189
FT  CARBOHYD 104 104
FT  CARBOHYD 144 144
SQ  SEQUENCE 255 AA; 28105 MW; 5C46DD59 CRC32;
-----
Query Match 2.3%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1.14e-00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;
-----
Db 25 DIVASVAQLG-VNLYQSYGSGQVSHFEDGDEFFYDLERKTVNOLPFRFR 78
QY 122 NLVNGTVPLGEMRGVAPSPDENSILVLEGGDEV-YSTIRKQYNGKIPRRIR 175
-----
RESULT 15
ID RBSR_ECOLI STANDARD; PRT; 329 AA.
AC P25551;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOSE OPERON REPRESSOR.
GN RBSR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6 AND 250-256.
RX STRAIN-K12;
RX MEDLINE: 93278299.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural and functional analyses of the repressor, RbsR, of the
RT ribose operon of Escherichia coli.";
RL PROTEIN SCI. 1:831-842(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE: 93315143.
RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL GENOMICS 16:551-561(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RL SUBMITTED (XXX-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SIMILARITY TO RIBOSE-BINDING PROTEINS.
RX MEDLINE: 93278300.
RA MAUZY C.A., HERMODSON M.A.;
RT "Structural homology between tbs repressor and ribose binding protein
RT implies functional similarity.";
RL PRONIN SCI. 1:843-849(1992).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RSDACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING
```

WAKELH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:52:40 1999; MasPar time 35.33 Seconds
Tabular output not generated.
860.513 Million cell updates/sec

Title: >US-09-041-236-2
Description: (12-568) from US09041236.pep (42 of 45)
Perfect Score: 4151
Sequence: 1 AQGHLRSGPRIFAVWKGHV.....ENVEQSCPEGHQSPNCILFI 557

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_Organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 48.560; Variance 74.668; scale 0.650

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
1	4151	100.0	666 4 075326 SEMAPHORIN L.	0.00e+00
2	2398	56.3	393 11 088371 SEMAPHORIN L. (FRAGMENT)	0.00e+00
3	1830	44.1	653 14 064906 SIMILAR TO GENBANK ACC	0.00e+00
4	737	17.8	748 11 062177 SEMAPHORIN A PRECURSOR	6.41e-155
5	725	17.5	751 11 062181 SEMAPHORIN E PRECURSOR	9.53e-152
6	716	17.2	749 4 013214 SEMAPHORIN V.	2.27e-149
7	716	17.2	750 4 093018 SEMAPHORIN V.	2.27e-149
8	710	17.1	751 4 099985 SEMAPHORIN E.	8.68e-148
9	705	17.0	751 13 042236 COLLAPSID 3.	1.81e-146
10	688	16.6	753 4 013372 SEMAPHORIN III FAMILY	5.44e-142
11	682	16.4	754 11 088633 SEMAPHORIN IV ISOFORM	2.06e-140
12	681	16.4	772 13 090607 COLLAPSID.	3.78e-140
13	682	16.4	785 11 088632 SEMAPHORIN IV ISOFORM	2.06e-140
14	681	16.4	785 4 013275 SEMAPHORIN IV.	3.78e-140
15	679	16.4	782 4 015704 SEMAPHORIN.	1.27e-139
16	669	16.1	772 11 062180 SEMAPHORIN D PRECURSOR	5.38e-137
17	664	16.0	772 11 063548 SEMAPHORIN III/COLLAPSID	1.11e-135
18	656	15.8	771 4 014563 SEMAPHORIN-III.	1.39e-133
19	618	14.9	782 11 062179 SEMAPHORIN C (SEM C)	1.22e-123
20	614	14.8	775 11 P70275 SEMAPHORIN H.	1.35e-122

21	599	14.4	861 11 009126 SEMAPHORIN J (SEMAPHOR	1.09e-118
22	592	14.3	775 4 015041 KIAA0331.	7.23e-117
23	572	13.8	785 13 042237 COLLAPSID 5.	1.13e-111
24	568	13.7	294 13 090664 COLLAPSID-3 (FRAGMENT)	1.23e-110
25	569	13.7	761 13 090663 COLLAPSID-2.	6.77e-111
26	538	13.0	862 4 092854 SEMAPHORIN.	7.00e-103
27	522	12.6	834 11 064151 SEMAPHORIN I (M-SEVA F	9.21e-99
28	511	12.3	299 13 090666 COLLAPSID-5 (FRAGMENT)	6.16e-96
29	473	11.4	295 13 090665 COLLAPSID-4 (FRAGMENT)	3.16e-86
30	474	11.4	760 11 062178 SEMAPHORIN B PRECURSOR	1.76e-83
31	461	11.1	730 5 026473 FASCICLIN IV.	3.54e-83
32	448	10.8	1074 4 013591 SEMAPHORIN F HOMOLOG.	6.95e-80
33	441	10.6	1077 11 062217 SEMAPHORIN F PRECURSOR	4.08e-78
34	416	10.0	1093 11 060519 SEMAPHORIN G PRECURSOR	7.94e-72
35	390	9.4	494 4 060408 SEMAPHORIN F (FRAGMENT	2.45e-65
36	392	9.4	888 11 035464 SEMAPHORIN VIA.	7.81e-66
37	378	9.1	712 5 026972 SEMAPHORIN-I PRECURSOR	2.32e-62
38	367	8.8	562 5 017330 CESEMA.	1.21e-59
39	367	8.8	771 5 024322 SEMAPHORIN-I.	1.21e-59
40	362	8.7	706 5 024323 SEMAPHORIN-II.	2.06e-58
41	341	8.2	284 11 054948 SEMAPHORIN N (SEMAPHOR	2.85e-53
42	316	7.6	886 11 054951 SEMAPHORIN Z.	3.24e-47
43	316	7.6	887 11 070141 SIMILAR TO SEMAPHORIN-	2.37e-39
44	283	6.8	770 5 044253 KIAA0463 PROTEIN (FRAG	2.37e-21
45	204	4.9	1963 4 075051	

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	666 AA.
ID	075326			
AC	075326;			
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L.			
GN	SEMA.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC	CATARRHINI; HOMINIDAE; HOMO.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 98389619.			
RA	LANGCE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;			
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA			
RT	viruses."			
RL	GENOMICS 51:340-350(1998).			
DR	EMBL; AF030698; G3523115; --			
SQ	SEQUENCE 666 AA; 74823 MM; F546B39D CRC32;			

Query Match 100.0%; Score 4151; DB 4; Length 666;				
Best Local Similarity 100.0%; Pred. No. 0.00e+00;				
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Db	44	AQGLRSGPRIFAVWKGHVQDRVDFGQTEPHTVLFHEPGSSVWVGGRKVKVLFDFPEG	103	
QY	12	AQGLRSGPRIFAVWKGHVQDRVDFGQTEPHTVLFHEPGSSVWVGGRKVKVLFDFPEG	71	
Db	104	KNASVTYVNIKSGKSLDKRDCENITLLRSEGLLAGCTNARHPSCNVLNGTVVPL	163	
QY	72	KNASVTYVNIKSGKSLDKRDCENITLLRSEGLLAGCTNARHPSCNVLNGTVVPL	131	
Db	164	GEMRGVAPSPDENSLVLFEGDEVSTIRKQEVNGKIPRFRIRGESELTSTVTMNPQ	223	
QY	132	GEMRGVAPSPDENSLVLFEGDEVSTIRKQEVNGKIPRFRIRGESELTSTVTMNPQ	191	
Db	224	FIKATIVHQDQYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGQGGESSLSVKWNT	283	
QY	192	FIKATIVHQDQYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGQGGESSLSVKWNT	251	
Db	284	FLKAMLVCSDAATNKNFNRLQDVFLLPDPGQWRDTRVYGVFNSPNWYSACVYSLGDID	343	
QY	252	FLKAMLVCSDAATNKNFNRLQDVFLLPDPGQWRDTRVYGVFNSPNWYSACVYSLGDID	311	

```
Db 344 KVERTSSLGKGYHSSSLNPRPGKCLPDQOPIPTTFQVADRHPEVAQVRPEMGPGLKTPLEH 403
|||||
QY 312 KVERTSSLGKGYHSSSLNPRPGKCLPDQOPIPTTFQVADRHPEVAQVRPEMGPGLKTPLEH 371
|||||
Db 404 SKYHYOKVAVHRMQASHGETFHVLYLTTDRGTTHKVVPEGEHSEHSAFNMIEIQPFRAA 463
|||||
QY 372 SKYHYOKVAVHRMQASHGETFHVLYLTTDRGTTHKVVPEGEHSEHSAFNMIEIQPFRAA 431
|||||
Db 464 ATQTMSLDAERKLYVSSQWESVSOVPLDICEVYGGCHGCLMSRDPYCGMDQGRCSIS 523
|||||
QY 432 ATQTMSLDAERKLYVSSQWESVSOVPLDICEVYGGCHGCLMSRDPYCGMDQGRCSIS 491
|||||
Db 524 SERSVLQISINPAEPHKECPNPKAPLQKVLAPNSRYLSCPMSRHATYSWRHKENV 583
|||||
QY 492 SERSVLQISINPAEPHKECPNPKAPLQKVLAPNSRYLSCPMSRHATYSWRHKENV 551
|||||
Db 584 EOSCEPGHSPNCILFI 500
|||||
QY 552 EOSCEPGHSPNCILFI 568
|||||

RESULT 2
ID Q88371 PRELIMINARY; PRT; 393 AA.
AC O88371;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN L (FRAGMENT).
GN SEMAL.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030699; G3523117; -.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43909 MW; 7C8B193C CRC32;

Query Match 56.3%; Score 2338; DB 11; Length 393;
Best Local Similarity 90.9%; Pred. No. 0.00e+00;
Matches 321; Conservative 14; Mismatches 14; Indels 4; Gaps 2;

Db 44 AQGHSRSGPRISAVWKG---QDHVDFSQPEPHTVLFHEPGSFVWVGGRGKYVHFNFPEG 100
|||||
QY 12 AQGHLRSRGPRIFAVWKGHVQDRVDFGQTEPHTVLFHEPGSSVWVGGRGKYVLFDFPEG 71
|||||
Db 101 KNASVTVNIGSTKGCQKQDCGNITILLRRGNGLLVCGTNARKPSCWNLVNSVMS 160
|||||
QY 72 KNASVTVNIGSTKGCCLDKDCENYITLLRRSEGLLACGTNARHPSCWNLVNGTVV-P 130
|||||
Db 161 LGEKGYAPFSPDENSLVLFEGDEVYSTIRKOEYNGKIPRFRIRGESELYTSDTVMQNP 220
|||||
QY 131 LGEKGYAPFSPDENSLVLFEGDEVYSTIRKOEYNGKIPRFRIRGESELYTSDTVMQNP 190
|||||
Db 221 QFIKATIVHQDQAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRDQDGGESSLSVSKWN 280
|||||
QY 191 QFIKATIVHQDQAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRDQDGGESSLSVSKWN 250
|||||
Db 281 TFLKAMLVCSDAATNRFNRLQDVFLLPDPSCQWRDTRVYGVSFPNPNVSAVCVYSLGDI 340
|||||
QY 251 TFLKAMLVCSDAATNRFNRLQDVFLLPDPSCQWRDTRVYGVSFPNPNVSAVCVYSLGDI 310
|||||
Db 341 DRVFTSSLUKGYHMLGNRPNGMCLPKKPOIPTTFQVADSHPEVAQVRPEMGP 393
|||||
QY 311 DKVFTSSLUKGYHSSLPNRPNGKCLPDQOPIPTTFQVADRHPEVAQVRPEMGP 363
|||||
```

```
RESULT 3
ID Q64906 PRELIMINARY; PRT; 653 AA.
AC Q64906;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALCELAHINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RX MEDLINE: 97201573.
RA ENSSER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RX MEDLINE: 97404659.
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C500;
RX ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U18243; G1000717; -.
DR EMBL; AF005370; G2337970; -.
SQ SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match 44.1%; Score 1830; DB 14; Length 653;
Best Local Similarity 47.0%; Pred. No. 0.00e+00;
Matches 264; Conservative 98; Mismatches 188; Indels 12; Gaps 12;

Db 67 AKSRIDKPRILVNLTDGFGQHRF-FGPOEHTVLFHSLNSDYYVGGNNTIYLFDFAH 125
|||||
QY 12 AQGHLRSRGPRIFAVWKGHVQDRVDFGQTEPHTVLFHEPGSSVWVGGRGKYVLFDFPEG 71
|||||
Db 126 SNASTALINITSHTNTHRLSTCENFITLLHNQTDGLLACGTNSQKPSW-LINNLTQF 184
|||||
QY 72 KNASVTVNIGSTKGCCLDKDCENYITLLRRSEGLLACGTNARHPSCWNLVNGTVV-P 130
|||||
Db 185 LQPKLGLAPFPSSGCLNVLFDNDYTSINLYKSLSGS-HKFRRIAGQVELYTSDTAMHR 243
|||||
QY 131 LGEKGYAPFSPDENSLVLFEGDEVYSTIRK-OEYNGKIPRFRIRGESELYTSDTVMQN 189
|||||
Db 244 PQFQVATVHKNESYDDKIYFFEQNSHSDFKOPHTVPRVGOVCSDDQGGESSLSVYKW 303
|||||
QY 190 PQFIKATIVHQDQAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRDQDGGESSLSVSKW 249
|||||
Db 304 TTFKLARLACYDYDTGRIYNEQLDIFIQAPENSWEEETLYGLFLSPWNFSACVYFTVKD 363
|||||
QY 250 NTEFLKAMLVCSDAATNKNFNRLQDVFLLPDPSCQWRDTRVYGVSFPNPNVSAVCVYSLGD 309
|||||
Db 364 IDHVFETSKLKNYHHKLTTPRPGQCKMKNHQVPTTFQVADRYPEVADPVYQKNNAMPEI 423
|||||
QY 310 IDKVFRTSLGKYHSSLPNRPNGKCLPDQOPIPTTFQVADRHPEVAQVRPEMGP 369
|||||
Db 424 IQSKYIYTKLIVRYEYG-GVFWATIFYLTITKGTIHYIYRYEDNSNTALNILEINPFO 482
|||||
QY 370 FHSKYHYOKVAVHRMQASHGETFH-VLYLTTDRGTTHKVVPEGEHSEHSAFNMIEIQPF 428
|||||
Db 483 KPAPIQNTLLDNTNLKLVNSENSEYSEVPLDCLSYGNDGCFSCFNRDPLCTWYNTC-S 541
|||||
QY 429 RAAAIQTMSLDAERKLYVSSQWESVSOVPLDICEVYGGCHGCLMSRDPYCGMDQGRCS 488
|||||
Db 542 -FK-QRVSVETGPGPANRTLSEMGDHYAPTVMKHOVSIPLLSNLSYLSCPVSNHADYFWT 599
|||||
QY 489 IYSSERSVLQISINPAEPH-KE-CPNPKPDKAPLQKVLAPNSRYLSCPMSRHATYSWR 546
|||||
```

Qy	411	GEQSHFAFIMEIQPERRAAAIQTMSLDARRKLVYSSQWVSVPLDLICEYVGSGCHG	470
Db	525	CCLARDPYCAWDGSACFRFOPTAKRRRRPRDRIKNGDPSTLCSGDSSHSLLEKKVLGVES	584
Qy	471	CLMSRDPYCGWDGRCISYISSERSVL--OSINPAEPHKCPKPKDKAPLQKVS LAPNS	528
Db	585	GSFALECEPRSLQAHVQW	602
Qy	529	RYYLSCPMESRHATYSW	545
RESULT 5			
ID	Q62181	PRELIMINARY; PRT; 751 AA.	
AC	Q62181;		
DT	01-NOV-1998	(TREMBREL. 08, CREATED)	
DT	01-NOV-1998	(TREMBREL. 08, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998	(TREMBREL. 08, LAST ANNOTATION UPDATE)	
DE	SEMAPHORIN E PRECURSOR (SEM E).		
GN	SEMAE OR SEME.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.		
RN	[]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NMRI;		
RX	MEDLINE; 95267431.		
RA	PUESCHEL A.W., ADAMS R.H., BETZ H.;		
RT	"Murine semaphorin D/collapsin is a member of a diverse gene family		
RT	and creates domains inhibitory for axonal extension.";		
RL	NEURON 14:941-948(1995).		
CC	-I- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO		
CC	SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.		
CC	-I- SUBCELLULAR LOCATION: EXTRACELLULAR (PROBABLE)		
CC	-I- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO.		
CC	MAXIMUM EXPRESSION BETWEEN DAYS 10-12 WITH MODERATE LEVELS FROM		
CC	DAY 13 UNTIL BIRTH.		
CC	-I- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.		
CC	CONTAINS ONE C2-LIKE DOMAIN.		
DR	ENBL; X85994; G854332;		
DR	MGD; MGI:107557; SEMAE.		
DR	PRAM; PF00047; ig; 1.		
KW	SIGNAL; IMMUNOGLOBULIN FOLD; MULTIGENE FAMILY; NEUROGENESIS;		
KW	DEVELOPMENTAL PROTEIN.		
FT	SIGNAL 1 20	POTENTIAL.	
FT	CHAIN 21 751	SEMAPHORIN E	
FT	DOMAIN 587 649	IG-LIKE C2-TYPE DOMAIN.	
SQ	SEQUENCE 751 AA; 85259 MW; B28D6CFE CRC32;		
Query Match 17.5%; Score 725; DB 11; Length 751;			
Best Local Similarity 33.0%; Pred. No. 9.53e-152;			
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28			
Db	167	SFNPNTVSVMIINEELFSWYI-DFMGTDAAIPLSLTKRMQLRTOHNSKWLSEPMFVD	225
Qy	139	PFSPDENSELFLFEGDEVYSIRKQYENGK-IPRFRIRGESELYTS--DTV-MONPOFIK	194
Db	226	AHVIPDGTDPNDKAVYFFERKLTDDNRSKTIQHSMIARICPDNTGORSI-VNKWTFEL	284
Qy	195	ATIVHQD-QAYDKIYFFREDPNPKNPEAPLVNSRVAQLCRGQGGESLSVKWNTFL	253
Db	285	KARLVCSVTDGDPETHFDELEVLL-ETDNP-RTLVYGIETSSVEKGSACVYHL	342
Qy	254	KAMLVCS-DA-ATNKFNRLQDVFLLPDPSGQWRDTRVYGVSNPNW-Y--SACVYSL	307
Db	343	SDIQTVPNGFAKHPGNHOLISYQGRIPYPRPGTCPGGAFTPMRTTKDPDQVTFIR	402
Qy	308	GDIDKVPRTS-SLK-G-----YHSSLPNRPKCLPDQ-QP-I-PTETF--QVADHRP	353
Db	403	NHPLMYNSISPIHRRPLIVRIGTDYKTKTAVDRVNAADG-RYHVLFLGTRGTQKVWV	461
Qy	354	EVAQRVPEMGL-KTPEL-H-SKIYQKVAVHRMQASHGETFHLVLLTDRGTIHKWVE	409
Db	462	LPTNSSAGSELILEEVFKNHVPIITMETSSKKQOLYSSNEGVSQVSLHRCHYGTAC	521

[illegible]

```
QY 195 ATIVHOD-QAYDDKIYFFREDNPNKNPEAPLNVSRVAOLCRGDQGGESSUSVSKWNTFL 253
Db 319 KARLVCSVPGEDGIETHFDELQDVFV-QOTQDV-RNPVIYAVFTSSGVSFGRSAVCVYSM 376
QY 254 KAMLVCS---DAATKNFNRLQDVFLLPDPGQWRDTRVYGVFSNP-WNY--SAVCVYSL 307
Db 377 ADIRMVFNPGFFAHKEGPNYQWMPFSGKMPYPRPGTCPGGTFPTPSMKSTKDYDDEVINPMR 436
QY 308 GDIDKVERTS-SLK-G--YH---SS-LPNRPRGKCLPDQ-QP-I-PTETF--QVAD--R 351
Db 437 SHPLMYQAVYPLQRRPLVVRT-GAPYRUTTIADVQVDACDG-RYEVLFGLTDRGTQVKVI 494
QY 352 -HPEVAQRVEPMG--PLKTPLFHSHYHOKVAVHRMQASHGETFHVLYLTTDRGTIHKVY 408
Db 495 VLPKDDQEMEELMEEVEVFKDPAPVKMTISSKRQQLYVASAVGVTLSLHRCQAYGAA 554
QY 409 E-PGEQHSFAFNEIOFFRAAAIQTMSLDAERRKLYVSSQWEVSVQVPLDLCEVYGG 467
Db 555 CADCLARDPYCAMDQACSRYTASSKRRSRQDYRHGNPIHQCRGFNSNANKNAVESVQ 614
QY 468 CHGCLMSRDPYCGWDQGRGIS-IYSSE-SVLQISINPAEPHKECP--NPKPKAPLQKVS 523
Db 615 YGVAGSAAFLEQPRSPQATYKW 637
QY 524 LA-PNSRYLSCPMESRHATYSW 545
```

Search completed: Thu Jul 8 19:53:49 1999
Job time : 69 secs.

W P S R E H (TW)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 8 19:58:34 1999; MasPar time 27.92 Seconds
Tabular output not generated. 427.348 Million cell updates/sec

Title: >US-09-041-236-2
Description: (41-601) from US09041236.pap (43 of 45)
Perfect Score: 4189
Sequence: 1 EPTVLFHEPGSSVWVGR.....EAQEGSYFRAQHQLLPED 561
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.297; Variance 153.226; scale 0.237

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	661	15.8	771 13	R71380	Human semaphorin III	4.91e-51
2	620	14.8	477 13	R74175	Human collapsin.	5.00e-47
3	608	14.5	775 33	W63748	Human semaphorin.	7.41e-46
4	606	14.5	861 22	W17658	Mouse CD100 antigen.	1.16e-45
5	599	14.3	861 32	W58540	Human semaphorin.	5.58e-45
6	544	13.0	862 22	W17657	Human CD100 antigen.	1.22e-39
7	509	12.2	776 32	W51313	Rat semaphorin W.	2.97e-36
8	463	11.1	441 13	R71381	Vaccinia virus semaph	7.86e-32
9	445	10.6	730 13	R71379	Grasshopper semaphori	4.15e-30
10	428	10.2	587 32	W51314	Human semaphorin W.	1.74e-28
11	390	9.3	974 33	W64221	Human secreted protei	7.01e-25
12	378	9.0	712 13	R71384	Tribolium semaphorin	9.51e-24
13	367	8.8	650 13	R71382	Drosophila semaphorin	1.03e-22
14	362	8.6	724 13	R71383	Drosophila semaphorin	3.04e-22
15	352	8.4	930 32	W57260	Human semaphorin Y.	2.63e-21
16	325	7.8	888 25	W19857	Human semaphorin Z.	8.68e-19

Rat semaphorin Z. 5.93e-18
Rat semaphorin Y. 4.15e-16
Variola major virus s 7.46e-04
Protein transcribed f 2.40e-01
Partial BRCA2 cancer 1.73e-01
Taq DNA polymerase I 2.83e-01
Taq DNA polymerase I 3.92e-01
Recombinant elastase. 8.75e-01
Sequence of nuclear r 8.75e-01
Taq DNA polymerase I 7.46e-01
Clas II S-receptor ki 7.46e-01
S receptor kinase pro 7.46e-01
A. mediterranei rifam 1.20e-02
BRCA2 cancer suscepti 1.20e-02
Partial BRCA2 cancer 1.20e-02
E. coli HSP (dnak). 1.02e-02
Mutant thermostable D 1.20e-02
Mutant thermostable D 1.20e-02
Mutant thermostable D 1.20e-02
DNA-polymerase REM-T3 1.20e-02
T. aquaticus DNA-poly 1.20e-02
Purified native therm 1.20e-02
Taq DNA polymerase I 1.20e-02
Taq DNA-polymerase RE 1.20e-02
Cleavase DN nuclease. 1.20e-02
Cleavase DA nuclease. 1.20e-02
Cleavase DV nuclease. 1.20e-02
Human breast cancer s 1.20e-02
Human breast and ovar 1.20e-02

ALIGNMENTS

RESULT 1

ID R71380 standard; Protein; 771 AA.
AC R71380;
DT 21-NOV-1995 (first entry)
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN W09507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; Page 60-63; 101pp; English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA;

Query Match 15.8%; Score 661; DB 13; Length 771;
Best Local Similarity 30.3%; Pred. No. 4.91e-51;
Matches 151; Conservative 126; Mismatches 174; Indels 47; Gaps 37;

Db 166 rgkspypdklltasllldgelysgta-adfmgdrdfairtllghhhprtcdhndrwndp 224

Qy	93	DCENYITLLER-RSEGLLAGCTNARHPSC--WNLVNGTVVPLGEM-RGVAPFSPDENS	148
Db	167	vmvggelysgts-ynflgsepiisrnnshsplrtetayalpwnepsfvfadviqspdgpe	225
Qy	149	LFEGDEVYSTIRKQOYNGKIPFRFRIRGESELYTSDTV--MQNPQFIKATIVHQ--D-Q	202
Db	226	geddkvffftesvveyefvfkImprvarcvkgdggllrtlg-kkwtlsflkarlicskp	284
Qy	203	AYDKKIYFFREDNPKNPAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKAMLVCSDA	262
Db	285	dsglvfnildqvfvlrap-gl-kepvfyavftpqlnnvglsavcaytlatveavfgrky	342
Qy	263	ATNKNENRLQDVFLLPDPSGOWRDRTRYGVFSNPWNY--SACVYVSLGDIKRVF-R--	315
Db	343	mqsatveqshkwyngpyvptprpaciadesaraanyteslnlpdktlqfvdhplmdd	402
Qy	316	--TSSL-----K-G-YHSSLNPNRPCKCL-PDQOPIP-TETFOVADRHPE-V-AQR-VE-	360
Db	403	svtpidnprplikdvnytoivdvrtqaldgtfydmfistdrgalhkvillkevhh-v	460
Qy	361	PMGPLKT-P-LFHSHYQKVAHVRMQASHGETFHLIYLTDRGTIHKVVPGEQSHSFA	418
Db	461	--leetglfdsepvlllsskkgrkfvyagsgvvgaplaefackhgs-cedcvlard	517
Qy	419	FNIMEIQPFRRAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDCEVYGGGCHGLMSRD	476
Db	518	pycawspaikacvtllhqeas	538
Qy	477	PYCGWDQG-R-CISIVSSERS	495
RESULT 5			
ID	W38540 standard; Protein; 861 AA.		
AC	W58540;		
DE	02-SRP-1998 (first entry)		
DE	Human semaphorin.		
KW	Human; semaphorin; diagnosis; nervous disease; immune disease;		
KW	nerve extension inhibiting activity.		
OS	Homo sapiens.		
PN	J10155490-A.		
PD	16-JUN-1998.		
PF	27-NOV-1996; 332900.		
PR	27-NOV-1996; JP-332900.		
PA	(SUMU) SUMITOMO SEIYAKU KK.		
DR	WPI; 98-391044/34.		
DR	N-PSDB; V31121.		
PT	New human semaforin gene - useful in the diagnosis of nervous system		
PT	and immune disorders		
PS	Claim 1; Page 10-12; 15pp; Japanese.		
CC	The present sequence represents human semaphorin (translated from the		
CC	Japanese specification as semaforin). Semaphorin has nerve extension		
CC	inhibiting activity. The semaphorin gene is useful for the diagnosis,		
CC	treatment and researches on nervous diseases and immune diseases.		
SQ	Sequence 861 AA;		
Query Match 14.3%; Score 599; DB 32; Length 861;			
Best Local Similarity 32.4%; Pred. No. 5,58e-45;			
Matches 143; Conservative 105; Mismatches 146; Indels 47; Gaps			
Db	107	eclnyirvlplstslslyvcgtnafgtdcdhnltsfkflgksedkgcrpdpahsyts	166
Qy	93	DCENYITLLER-RSEGLLAGCTNARHPSC--WNLVNGTVVPLGEM-RGVAPFSPDENS	148
Db	167	vmvggelysgts-ynflgsepiisrnnshsplrtetayalpwnepsfvfadviqspdgpe	225
Qy	149	LFEGDEVYSTIRKQOYNGKIPFRFRIRGESELYTSDTV--MQNPQFIKATIVHQ--D-Q	202
Db	226	geddkvffftesvveyefvfkImprvarcvkgdggllrtlg-kkwtlsflkarlicskp	284
Qy	203	AYDKKIYFFREDNPKNPAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKAMLVCSDA	262
Db	285	dsglvfnildqvfvlrap-gl-kepvfyavftpqlnnvglsavcaytlatveavfgrky	342

QY 263 ATNKNFRLQDVFLLPDPGQWRTRVYGFSPNPNY---SACVYSLGIDIKVF-R--- 315
Db 343 mqsatveqstkwryngpytpgpcidsearaanytsslnlpdktlqfvkdhplmdd 402
QY 316 --TSSL-----K-G-YHSSLPNPRGKCL-PDQQPIP-TETFOVADRPE-V-AQR-VE- 360
Db 403 svtpidnrplikkdvnyqtgldvtdqldgtfydvnmfistdrgalhhkavlltkevh--v 460
QY 361 PMGPKLT-P-LFHSKYHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVPEGQEHSEA 418
Db 461 --ieetqlfdfevplclllsskgrfvyagsngvvqaplafcekhgs-ceedcvlard 517
QY 419 FNIMEIQPFRAAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDLCEVYGGCHGLMSRD 476
Db 518 pycawspaikacvtlhqeeas 538
QY 477 PYCGWDOG-R-CISIIYSERS 495

RESULT

ID W17657 standard; Protein; 862 AA.
AC W17657:
DT 24-JUL-1997 (first entry)
DE Human CD100 antigen.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..41
FT /label_ Sig_peptide 42..862
FT /label_ Mat_protein 42..553
FT /label_ Semaphorin_domain 554..630
FT /label_ Ig-like_domain 631..733
FT /label_ Stalk_domain 735..752
FT /label_ Transmembrane_domain 753..862
FT /label_ Cytoplasmic_domain 808..815
FT modified_site 808..815
FT /label_ Phosphorylation
FT /note_ "putative tyrosine phosphorylation site"
PN W09717368-A1.
PD 13-MAY-1997.
PF 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND) DANA FARBER CANCER INST.
PI Boussioutis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB: T60665.
DR Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Claim 7: Page 70-72: 135pp: English.
CC Human CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60665) isolated by COS cell expression cloning. CD100
CC polypeptides and fusion proteins, nucleic acids, and host cells
CC expressing CD100 can be utilised in diagnostic and therapeutic
CC methods involving modulation of B and T cell responses, neuron
CC axonal growth and immune cell-nerve cell interaction.
SQ Sequence 862 AA;

Query Match 13.0%; Score 544; DB 22; Length 862;
Best Local Similarity 29.3%; Pred. No. 1.23e-39;
Matches 137; Conservative 134; Mismatches 189; Indels 55; Gaps 39;

Db 107 ecinyirvqlpsatslyvcgtnafqpadhlnitsfkflgknedkgrocpfpahsyts 166
QY 93 DCENYITLLRRSE-GLLAGCTNARHPSCWNLVNGTVVPLG--EM-RGYAPFSPDEN-SL 147
Db 167 vmvdg-elysgts-ynflgsepiisrnsnshsplrteyapwlnepsfvfadvirkspdsp 224
QY 148 VLFEGDEVYTIKQYNGKIPRFRIRGESELYTSDTV--MQNPQIKATIIHQ--D-- 201
Db 225 dgeddvvyffftvsveyefvflprialrvckgqggglrtllq-kkwtstflkarlicsr 283
QY 202 QAYDDKIYFFREDNPDKNPEALNVSRVAQLCRGQGGESSLSVKWNTFLKAMLVCS 261
Db 284 pdsghlfnvldrvfvlrslp-gl-kvpvfyalfpqlnnvglsvacaynlnstaeefshgk 341
QY 262 AATNKNFRLQDVFLLPDPGQWRTRVYGFSPNPNY---SACVYSLGIDIKVF--R 315
Db 342 ymgsttveqstkwryngpytpgpcidsearaanytsslnlpdktlqfvkdhplmdd 401
QY 316 ----TS---S-LK-G-YHSSLPNPRGKCL-PDQQPIP-TETFOVADRPE-V-AQR-VE 360
Db 402 dsvtpidnrplikkdvnyqtgldvtdqldgtvydvnmfistdrgalhhkai--s-lehav 458
QY 361 -PMGPKLT-P-LFHSKYHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVPEGQEH 417
Db 459 hi-ieetqlfdfevplclllsskgrfvyagsngvvqaplafcekhgt-ceedcvlar 516
QY 418 AFNIMEIQPFRAAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDLCEVYGGCHGLMSR 475
Db 517 dpycawspptatcvlqhatesprglieqmsgdasvcpd-ks-kgsyrqhfhhgtael 574
QY 476 DPYCGWDOG--GRCISIIYSERSVLOSINP-AEPHKECPNPKAPLQKVS LAPNSRYL 532
Db 575 kcsqksnlarvfwkfngvllkaespyglmgrknllif-nlsegdsgvqvclsee 628
QY 533 SCPMESRHATYSWRHKNVENQSCPE--GHQSPNCILFIENLTAQQYGHVFCEAOE 585

RESULT

ID W51313 standard; Protein; 776 AA.
AC W51313:
DT 08-SEP-1998 (first entry)
DE Rat semaphorin W.
KW Rat; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW Immunosuppressant; gene therapy; diagnosis; research reagent.
OS Rattus norvegicus.
PN W09815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMO) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-PSDB: V07279, V07280.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Claim 1: Page 60-64: 90pp: Japanese.
CC The present sequence represents rat semaphorin W semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 776 AA;

Query Match 12.2%; Score 509; DB 32; Length 776;
Best Local Similarity 27.3%; Pred. No. 2.97e-36;
Matches 150; Conservative 135; Mismatches 212; Indels 53; Gaps 37;

Db 71 sallvdpsathtlyvgardsfaltlbfsgprpridwmpv-ethrqncrkkgkdechn 129
QY 44 TVLFHEPGSSVWVGGRGVY-L-FDFPEGKNASV-RTVNIIGTKGSL--DKR-D-CEN 96


```
QY 148 VLFGEDEVYSIRKOEYNGKIPRFRIRGESELYTSVTMQNPQFIKATIIVHQDAYDDK 207
Db 218 iffretaveyincgaiky-srvarckhdkg-phqggdrwtstfiksrlncsvpgdyp 275
QY 208 IYFFREDNPD-KNPEAPLNVSRVAQLCRGQGESSLSVSKWNTFLKAMLVCSDAATNK 266
Db 276 fyfneigstsdiegnvggq-veklygvfttpvnsiggsavcafsmksilesfdgfkfe 334
QY 267 -NFENRLQDVF-LLPDP-SGQWRDRVYGVFSNPWN-Y--SAVCVYSL-G--D-IDKVFRT 316
Db 335 qetmnsnwlavpslkvpeprpgqcvndsrtpdvsvnfvkshtlmdavpafft-rpili 393
QY 317 -SSLKG-YHS--SL--PNPRFGKLPDQPIPTETTFQVADRHPVQAQRVEPMGLKPTLF 370
Db 394 rislqyftklavdgqrvtpdgkaydlfigtdgdkvkalnsasfdssdtdvdsvleel 453
QY 371 H-S-KYHYQKAV-HRMOASGEHFHVLTLTDRTGTHKVVVEPG--EOEHSFAFNIME-I 424
Db 454 qvlppgvvknlyvrmdgdsklvvvsddellaiklhrgsdkitncrecvslqdpvca 513
QY 425 QPFRRAAAIOTM---SLDAERRKLYVSSQWVEVSQVPLDLCEVYG-GGCHGLMSRDPYCG 480
Db 514 wdnvelkctavgsdpws 530
QY 481 WDG-G-RCISIISSERS 495

RESULT 10
ID W51314 standard; Protein; 587 AA.
AC W51314;
DT 08-SEP-1998 (first entry)
DE Human semaphorin W.
KW Human; semaphorin W.
KW Immunosuppressant; nerve extension inhibitor; antiallergic; anticancer;
OS Homo sapiens.
PN W09815628-A1.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMU ) SUMITOMO PHARM CO LTD.
PI Kkuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-PSDB: V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 10.2%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 1.74e-28;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgdgdeylffttetsrafdsyerkivprvarvcagdlggkrtliq-grwtffikadilcp 115
QY 201 DOAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGQGESSLSVSKWNTFLKAMLVCS 260
Db 116 gpehgrassvlqdvavrlpelga-gtpifygfssqwegatisavcafrpqdirtvlnpg 174
QY 261 DAAATNKNFNRLQDVFLLPDPDSGQWRDRVYGVFSNPWN---YSAVCVYSLGDDIKVPTS 317
Db 175 frelhdcnrglvpvndvpqprgceitnmklrhfgsslsipdrvlrtfirdhplmdrp 234
QY 318 --SUK-GYHSSLP---N----PRPGKCLPQQPIP--TETTFQVADRHPV-V-AQR-VE-P 361
Db 235 vfpadghpllvtttdtaylrvvahrvtslsgkeydvlylgtedghlhravigaq-ls-vl 292
```

```
QY 362 MGPKLT-PLFHSK-YHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVVEPGEQHSFAF 419
Db 293 edial--fpeqpqvemkl-yhsw-llvgqrvtevtqvtntncgrlqs-csecilaqdpvc 347
QY 420 NIMEIOPFRRAAAIOTMSLDAERRKLYVSSQWVEVSQVPLDLCEVYGGCHGLMSRDPYC 479
Db 348 awsfriecvahagehrglvgldesadvslcpkegerpvvfepvataahvvlpcsp 407
QY 480 GWD-Q-GRCLISIISSERSVLQSIINPAEPHKECPNRPDKAPLOKVSUAPNSRYLSCPME 537
Db 408 savascvwthqpsgv 421
QY 538 SRHATYSWRHKENV 551

RESULT 11
ID W64221 standard; Protein; 974 AA.
AC W64221;
DT 06-OCT-1998 (first entry)
DE Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN W09827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GEMI ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavalie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-362774/31.
DR N-PSDB: V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17j; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 9.3%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 7.01e-25;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechfivkllknddalfvcgtfnafnpscrnyknmdtlepfgfdefsgmarcpydakhn 172
QY 91 KRDCENYIT-LLERRSEGLLACCTNARHPSCWNLVNGTVVPLG-EMRGVA--PFSPDENS 146
Db 173 valfadgklysatvt-dflaidavirsl-geep--tlrtvkhdkswlkepyfvqadvy 228
QY 147 LVLFGEDEVYSIRKOEYNG-KIPRFRIRGESELYTSVTMQNPQFIKATIIVHQDAYD 205
Db 229 dylyffirelavyemkvvfrvaqvcndngsqgrvlekqvtflkarlncsvpgds 288
QY 206 DKIIYFFREDNPDKNPEAPLNVSRVAQLCRGQGESSLSVSKWNTFLKAMLVCSDAATN 265
Db 289 hfyfnllqavtdvirng--rdv-vlatfstpynslpgsavcaydmldiasvftgrfkeq 345
QY 266 K-NFNRLQDVFLLPDPDSGQWRDRVYGVFSNPWN-Y--SAVCVYSLGDDIKVF--RTSSL 319
```


Oy 424 -IQPF---R-----RAA--A--IQTMSLDAERRKLIYSSQWEVSQVPLDLCEVYGGCH-G 470
Db 529 clasqdpycgwhsrgcvdrlrgsgtdvdqagngqesmehgdcqd 572
Oy 471 CLMSRDPYCGWDQGR-CISIYSSE-RSVLQSIINPAE-PHKECPN 511

Search completed: Thu Jul 8 19:59:11 1999
Job time : 37 secs.

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

mpsrch pp protein - protein database search. using Smith-Waterman algorithm

```
Run on: Thu Jul 8 19:57:44 1999; MasPar time 25.55 Seconds
879.945 Million cell updates/sec
Tabular output not generated.
```

```
>US-09-041-236-2
Description: (41-601) from US09041236.pep (43 of 45)
Perfect Score: 4189
Sequence: 1 EPTVLFHEPGSSGVVVGRR.....EAGSGSYFRAOHWLLPED 561
```

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:
pir60
1:pir1 2:pir2 3:pir3 4:pir4
```

```
Statistics:      Mean 48.952:  Variance 86.379:  scale 0.567
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Query		DB	ID	Description	Pred. No.
	Match	Score				
1	43.6	1825	2	T03102	semaphorin homolog A3	0.00e+00
2	37.7	737	2	I48744	semaphorin A - mouse	1.40e-13
3	17.3	725	2	I48748	semaphorin E - mouse	8.12e-13
4	17.1	749	2	G01856	semaphorin V - human	9.55e-13
5	16.4	718	2	G02173	semaphorin III family	2.57e-12
6	16.3	681	2	A49069	collapsin - chicken	1.03e-12
7	16.0	669	2	I48747	semaphorin D - mouse	5.77e-11
8	15.9	667	2	I58169	semaphorin III - mous	1.66e-11
9	15.8	771	2	D49423	semaphorin III precur	3.90e-11
10	14.8	782	2	I48746	semaphorin C - mouse	2.49e-10
11	12.5	834	2	S66498	M-sema F protein prec	1.17e-85
12	11.5	760	2	I48745	semaphorin B - mouse	2.73e-76
13	11.2	403	2	E42521	A39r protein - vaccin	4.50e-74
14	11.1	463	2	S29912	hypothetical protein	1.59e-72
15	10.7	730	2	J07938	fasciclin IV precursor	5.42e-69
16	10.7	1074	2	JC5928	semaphorin F precursor	3.26e-69
17	9.3	711	2	A49423	semaphorin I precursor	2.53e-56
18	8.8	655	2	B49423	semaphorin I - fruit	1.40e-51
19	8.6	724	2	C49423	semaphorin II precurs	1.65e-50
20	7.0	293	2	JQ1775	SaLL9r protein - vacc	5.74e-36
21	4.7	1884	2	JC4975	plexin 2 precursor -	1.01e-16
22	4.1	1872	2	AC9766	plexin 3 precursor -	5.92e-12
23	4.1	1905	2	I51533	plexin - African claw	3.92e-12

QY 398 TDRGTHKVVPEGEHSEFAFNIMEIQFRRAAAIQTMSLDAERKLYVSQWSEVSYQP 457
Db 512 LDLCVYGNDFCFSCMSRDLCTWYNNTC-S-FK-QRVSVETGGPANRTLSEMCGDHYAP 568
QY 458 LDLCVYGGCHCLMSRDPYCGWDGRCISYISRSRVSQSLNPAEPH-KE-CPNPKPD 515
Db 569 TVYKHQVSIPLJNSLYSLSPAVSNHADYFTWKDFTEKRCRCHVTKHNDICILLIANSTTAT 628
QY 516 KAPLQKVS LAPNSRYLYLSCPMESRHATYSWRHKNENVEQSCPEHQHQPNCILFIENLTAAQ 575
Db 629 NGTHVCNMKEDS 640
QY 576 YGHYCEAQEGS 587

RESULT 2
ENTRY semaphorin A - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
DATE 04-Sep-1998

ACCESSIONS I48744
REFERENCE Puschel, A.W.; Adams, R.H.; Betz, H.
#authors Neuron (1995) 14:941-948
#journal Murine semaphorin D/collapsin is a member of a diverse gene
#title family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48744

GENETICS
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-748 ##label RES
##cross-references EMBL:X85990; NID:g854323; PID:g854324

CLASSIFICATION semA
#superfamily semaphorin
#length 748 #molecular-weight 82894 #checksum 9017

Query Match 17.6%; Score 737; DB 2; Length 748;
Best Local Similarity 33.7%; Pred. No. 1,40e-134;
Matches 168; Conservative 112; Mismatches 166; Indels 52; Gaps 36;

Db 112 ECMNFVRLHAYNHTLLACRTGAFHPTCALMRWATAGTSTGPEKLEDKGKTPYDP 171
QY 93 DCENYITLLERRSEG-LLACGTNARHPSC- ---WNLVNGT-V-P--LGERMGVAPESP 142
Db 172 RHPPSVLYGEELYSV-TADLMGRDFTIFRSLGQNPSLRTEPHDSRWLNPEKFKVFWI 230
QY 143 DENSLVLFEGDEVYSTIRKOEYNGK-IPRFRIRGESELYT-S-DTV-MONPQFIKAT-I 197
Db 231 PESENDDDKIYFFFRSAVEAAPAMGRMSVSRVGOICRNDLGGORSL-VNKNWTFLLKAR 289
QY 198 VHODQAYDDKIYFFREDNPKNPE-APLNVSRVAQLCRGDOGGESSLSVSKWNTFLKAM 256
Db 290 LVCSVPGVEGDTHFDOLQDVFLLS-SR-DRQTPLLYAVESTSGVFGQSAVCVYSNNDVR 347
QY 257 LVCS--DAATNKNFNRLQDVFLLPDPSGQWRDTRVYGVFSPNPN-Y--SACVCVSLGDD 311
Db 348 RAFLGPLPHKEGPTHQWVYQGRVYPRPGMCPSTKFTGTSKDPDPPDVIOFGRNHPLM 407
QY 312 KVFTS-SLK-G-----YHSSLNPRPGKLPDQ-QPIP-TETF--QVAD--R-HPEV 355
Db 408 YNPVLPMSGG-R-PLFLQVAGYTFFTQIADRVAADGH-YDVLFTGTDVGTVLKVISVPK 464
QY 356 AQRVPMGPKTPLF-H--SKYHOKVAVHRMOASHGETFHVLYLTDRGTTHKVVY-P- 410
Db 465 GRPNSEGLLLEQLQVDFESAITSQKISKRQQLYASRAAQAIALHRCRTALGRACAE 524
QY 411 GGEHSEFAFNIMEIQFRRAAAIQTMSLDAERKLYVSQWSEVSYQVPLDLCEVYGGCHG 470
Db 525 CCLARDPYCANDGSACTRFOPTAKRRFRQDITRNGDPSTLCSDGSSHSLVLEKKYLVGES 584

QY 471 CLMSRDPYCGWDGRCISYISRSVSL--QSINPAEPHKECPNPKDPKAPLQKVS LAPNS 528
Db 585 GSAFLECEPRSLQAHVQW 602
QY 529 RYILSCPMESRHATYSW 545

RESULT 3
ENTRY semaphorin E - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
DATE 04-Sep-1998

ACCESSIONS I48748
REFERENCE Puschel, A.W.; Adams, R.H.; Betz, H.
#authors Neuron (1995) 14:941-948
#journal Murine semaphorin D/collapsin is a member of a diverse gene
#title family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48748

GENETICS
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-751 ##label RES
##cross-references EMBL:X85994; NID:g854331; PID:g854332

CLASSIFICATION semE
#superfamily semaphorin
#length 751 #molecular-weight 85259 #checksum 8961

Query Match 17.3%; Score 725; DB 2; Length 751;
Best Local Similarity 33.0%; Pred. No. 8.12e-132;
Matches 148; Conservative 110; Mismatches 149; Indels 41; Gaps 28;

Db 167 SFNPVNTVSVNINEELFSGMYI-DFMGTDAAIFRSLTKRQMLRTDQNSKWLSEPMFVD 225
QY 139 PFPSPDENSLLVFEDEVYSTIRKOEYNGK-IPRFRIRGESELYTS--DTV-MONPQFIK 194
Db 226 AHVIPDGTDPNDAKVYFFFKERLTDNNRSTKQIHSMIARICPNDTGGORSL-VNKNWTFLL 284
QY 195 ATIVHQD-QAYDDKIYFFREDNPKNPEAPLNVSRVAQLCRGDOGGESSLSVSKWNTFL 253
Db 285 KARLVCSVDEDDGPEHDELEDFELL-ETDNP-RTLLVYGIFFTSSSVFKGSAYCVYHL 342
QY 254 KAMLVCS--DA-ATNKNFNRLQDVFLLPDPSGQWRDTRVYGVFSPNPN-Y--SACVCVSL 307
Db 343 SDIQTVFNGPFAHKEGPNHQLISYQGRIPYPRPGTCPGAFTPNMRTTKDPPDDVVVTFIR 402
QY 308 GDIDKVFTS-SLK-G-----YHSSLNPRPGKLPDQ-QP-I-PTETF--QVADRHP 353
Db 403 NHPLMYSNISPIHRRPLVIRIGTDYKTKIAVDVNAADG-RYHVLFLGTDRGTGVKVVV 461
QY 354 EVAQRVPMGPKTPLF-H--SKYHOKVAVHRMOASHGETFHVLYLTDRGTTHKVVY 409
Db 462 LPTNSSASGELILELEVFKNHVPIITTEISSKKQQLYSSNEGVSQVSLHRCRHIYTAG 521
QY 410 -PGEHSEFAFNIMEIQFRRAAAIQTMSLDAERKLYVSQWSEVSYQVPLDLCEVYGGC 468
Db 522 ADCCCLARDPYCANDGSCSRFPYPTGKRKRQDVRHGNPLTCGRGFNLKATRNAAEIVQY 581
QY 469 HGCLMSRDPYCGWDGRCISYISSE--RSVLQSLNPAEPHKECP--NPKPKAPLQKVS 524
Db 582 GVRNNTFLECAPKSPQASIKWLLQDK 609
QY 525 A-PNSRYLSCPMESRHATYSWR-HKEN 550

RESULT 4
ENTRY semaphorin V - human
TITLE #formal_name Homo sapiens #common_name man
ORGANISM 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
DATE 04-Sep-1998

[illegible]

Db	470	HDLEVLLEEMTWVREPTTISAMELSTKQOOLYIGSTAGVAQLPHRCDIYGKACAECLM	529
Qy	415	HSFAFNIME-IQPFRAAAIQTNSLDAERKLYVSSQWEVSQVPLDLCEYVGCGCHCLM	473
Db	530	ARDPYCAWDGSSCSRYFPTAKRRTRRDIRNGDPLTHCSDLHDHNNHGSLSLERIIYG	589
Qy	474	SRDPYCGWDGRCIGSIYSSE-RSVL-QSINAPBPHKECPN-PPDK--AP-LO-KVSLA-	525
Db	590	ENSTTFLECSPKSQORALVY-WQ 610	
Qy	526	PNSRYILSC-PMESRHATYSWR 546	
RESULT	8		
ENTRY		I58169 #type fragment	
TITLE		semaphorin III - mouse (fragment)	
ORGANISM		#formal name Mus musculus #common name house mouse	
DATE		26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change	
ACCESSIONS		04-Sep-1998	
REFERENCE		I58169	
#authors		Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;	
#journal		Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.	
#title		Neuron (1995) 14:949-959	
#cross-references		Semaphorin III can function as a selective chemorepellent	
#accession		pattern sensory projections in the spinal cord.	
		MUID:95267432	
		I58169	
		#status preliminary; translated from GB/EMBL/DBJ	
		#molecule_type mRNA	
		#residues 1-666 #label RES	
		#cross-references GB:L40484; NID:g703189; PID:g703190	
GENETICS			
#gene		SemaIII	
CLASSIFICATION		#superfamily semaphorin	
SUMMARY		#length 666 #checksum 9654	
Query Match		15.9%; Score 667; DB 2; Length 666;	
Best Local Similarity		30.1%; Pred. No. 1.66e-118;	
Matches		150; Conservative 130; Mismatches 168; Indels 50; Gaps	
Db	60	RGKSPYDKLLTASLLIDGELYSCTA-ANFMRDFAIFRTLGHHPITRTEQHSRWLNDP	118
Qy	135	RGVAPFDENSLVFGDEGYSTIRKQYNGK-IPFRRIERGESELYTS--DTV-MQNP	190
Db	119	RFISAHLIPESDNPEDDKVYFFFRENAIDGESHGKATHARIGQICKNDFGGRSL-VNKK	177
Qy	191	QFIKATIVHQ-DQAYDDKIYFFREDNPDKNPAPLNVSRVAQLCRGDQGESLSVSKW	249
Db	178	TTFLKALICSVPGNPIDITHDELQVFLM-NSKDP-KNPVIYGVFTTSSNIFKGSVAC	235
Qy	250	NTEFLKAMLVCS--DA-ATNKNFNRLQDVFLLPDPGOWRDRYGVFSPNWN-Y--SVC	303
Db	236	MYMSDVRVRLLYGPAHRDGNQWYVPGYQGVYPRPGTCPSKTFGGFSDSTKDLDDVIT	295
Qy	304	YISLGDIDKV----F--RTSS-LK--GYHSLSNPRPKCLPQO-QPI-ETETF--QV-A	349
Db	296	FARSHPAMYNPVFPINN-RPIMIKTDVNYQTOIVVDVDAEDGQ-YDVNFICTDVGTVL	353
Qy	350	-DR-HPEVAQRVEPMGLKTPLFHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTIH	405
Db	354	KVSVVPKETHDLEEVLEEMTWVREPTTISAMELSTKQOOLYIGSTAGVAQLPHRCDI	413
Qy	406	KWVE-PGEQESHSAFNIME-IQPFRAAAIQTNSLDAERKLYVSSQWEVSQVPLDLCEV	463
Db	414	YGKACAECLLARDPYCAWDGSSCSRYFPTAKRRTRRDIRNGDPLTHCSDLQHDHNNHP	473
Qy	464	YGGCHCCLLSRDPYCGWDGRCISIIYSSE-RSVL-QSINAPBPHKECPN-PPDK--AP	518
Db	474	SLEERIIVGVENSTFLECSPKSQORALVYVQFORNEDRKEEIKMGDHIIRTEQGLLRS	533
Qy	519	-LO-KVSLA-PNSRYILSC-PMESRHATY-SW-RKHENVEOSCEPG-HO-SPNCILFIEN	570

[illegible]

THIS PAGE BLANK (USPTO)

W A R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:56:04 1999; MasPar time 17.60 Seconds
Tabular output not generated. 900.921 Million cell updates/sec

Title: >US-09-041-236-2
Description: (41-601) from US09041236.pep (43 of 45)
Perfect Score: 4189
Sequence: 1 EPTHTLVFHEPGSSVWVGGR.....EAQEGSVFRAQHWQLLPED 561

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 50.026; Variance 75.715; scale 0.661

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	470	11.2	VA39_VACC	PROTEIN A39.	2.89e-86
2	463	11.1	VA39_VACCV	PROTEIN A39.	1.80e-84
3	163	3.9	SEX_HUMAN	TRANSMEMBRANE PROTEIN H	8.11e-13
4	105	2.5	US08_HCMVA	HYPOTHETICAL PROTEIN H	3.46e-02
5	102	2.4	NRT2_RAT	T-CELL ECTO-ADP-RIBOSY	1.01e-01
6	98	2.4	NRT1_RAT	T-CELL ECTO-ADP-RIBOSY	2.90e-01
7	99	2.4	EF1A_TRIRE	ELONGATION FACTOR 1-AL	2.90e-01
8	99	2.4	VRP2_SALTY	65 KD VIRULENCE PROTEI	2.90e-01
9	99	2.4	VRP2_SALCH	65 KD VIRULENCE PROTEI	2.90e-01
10	99	2.4	VRP2_SALEN	65 KD VIRULENCE PROTEI	2.90e-01
11	99	2.4	VRP2_SALDU	65 KD VIRULENCE PROTEI	2.90e-01
12	99	2.4	PMAL_AJECA	PLASMA MEMBRANE ATPASE	2.90e-01
13	95	2.3	HA21_HUMAN	HLA CLASS II HISTOCOMP	1.13e+00
14	95	2.3	VL2_HPV5B	MINOR CAPSID PROTEIN L	1.13e+00
15	95	2.3	VL2_HPV05	MINOR CAPSID PROTEIN L	1.13e+00
16	95	2.3	AMOH_ARTGO	HISTAMINE OXIDASE (EC	1.13e+00
17	98	2.3	DNBI_HSV6U	MAJOR DNA-BINDING PROT	4.09e-01
18	95	2.3	MEI_HSV6U	HEPATOCYTE GROWTH FACT	1.13e+00
19	94	2.2	RBR_ECOCI	RIBOSE OPERON REPRESSO	1.57e+00
20	91	2.2	DDL_ENTFA	D-ALANINE--D-ALANINE L	4.16e+00
21	91	2.2	VP39_NPVLD	MAJOR CAPSID PROTEIN.	4.16e+00
22	93	2.2	DP38_MYCPN	DNA POLYMERASE III, BE	2.18e+00
23	92	2.2	CEM1_YEAST	3-OXOACYL-(ACYL-CARRIE	3.02e+00

24	92	2.2	460	1	EF1A_NEUCR	ELONGATION FACTOR 1-AL	3.02e+00
25	91	2.2	498	1	ELAS_FSEAE	PSEUDOLISIN PRECURSOR	4.16e+00
26	91	2.2	598	1	NOT_HUMAN	IMMEDIATE-EARLY RESPON	4.16e+00
27	91	2.2	707	1	GCVK_HCMVA	GANCICLOVIR KINASE (EC	4.16e+00
28	94	2.2	947	1	PLA2_YEAST	PLASMA MEMBRANE ATPASE	1.57e+00
29	94	2.2	982	1	POL_HTLV2	POL POLYPROTEIN [CONTA	1.57e+00
30	93	2.2	1029	1	END1_YEAST	VACUOLAR BIOGENESIS PR	2.18e+00
31	94	2.2	1069	1	ENTK_MOUSE	ENTEROPEPTIDASE (EC 3.	1.57e+00
32	94	2.2	1115	1	IRE1_YEAST	SERINE/THREONINE-PROTE	1.57e+00
33	92	2.2	1132	1	DNBI_HSV62	MAJOR DNA-BINDING PROT	3.02e+00
34	94	2.2	1200	1	DXH8_CAEEL	PUTATIVE PRE-MRNA SPLI	1.57e+00
35	92	2.2	1663	1	CO3_MOUSE	COMPLEMENT C3 PRECURSO	3.02e+00
36	91	2.2	1666	1	CO3_CAVPO	COMPLEMENT C3 PRECURSO	4.16e+00
37	90	2.1	119	1	RL19_MYCPN	50S RIBOSOMAL PROTEIN	5.72e+00
38	90	2.1	213	1	AMEX_BOVIN	AMELOGENIN, CLASS I PR	5.72e+00
39	90	2.1	254	1	HA22_HUMAN	HLA CLASS II HISTOCOMP	5.72e+00
40	89	2.1	404	1	NIFS_ECOCI	NIFS PROTEIN HOMOLOG.	7.83e+00
41	89	2.1	437	1	RFBB_MVXXA	O-ANTIGEN EXPORT SYSTE	7.83e+00
42	89	2.1	518	1	VL2_HPV47	MINOR CAPSID PROTEIN L	7.83e+00
43	89	2.1	649	1	RA32_SCHPO	DNA REPAIR PROTEIN RAD	7.83e+00
44	90	2.1	1874	1	POUR_XYMVJ	RNA REPLICASE POLYPROT	5.72e+00
45	89	2.1	3418	1	BRC2_HUMAN	BREAST CANCER TYPE 2 S	7.83e+00

ALIGNMENTS

RESULT 1
ID VA39_VACC STANDARD; PRT; 403 AA.
AC P21062;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE PROTEIN A39.
GN A39R.
OS VACCINIA VIRUS (STRAIN COPENHAGEN).
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91021027.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RT "The complete DNA sequence of vaccinia virus."
RL VIROLOGY 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RL VIROLOGY 179:517-563(1990).
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M35027; G335517; -

DR PIR; EA2521; E42521.

SQ SEQUENCE 403 AA; 45741 MW; FE0DB9AB CRC32;

Query Match 11.2%; Score 470; DB 1; Length 403;

Best Local Similarity 33.5%; Pred. No. 2.85e-86;

Matches 86; Conservative 59; Mismatches 95; Indels 17; Gaps 16;

Db 77 LVCGTNNNPKCWK-IDGSDDPKHKRGYAPYQNSKVITISYN-ECVLSIDINISK-EG-I 132

Qy 109 LACGTNARHPSCWNLVNGTVVPLGEMRGYAPFSPDENSELVLFEGDEVTSTKQYNGKI 168

Db 133 KWRREFDCPGYDLTADNPVTKDG-LRGAFAVDKGTVD-KVYILFTDTIGSKR--I-VK 187

Qy 169 PRFRIRGE-S-ELYTSDTVNQNPQFIKATIVHQDAYDDKIYFFREDNPDPKPEAPLN 226

OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
[1] SEQUENCE FROM N.A.
RN MEDLINE; 90138997.
RX KOCH F., HAAG F., KASHAN A., THIELE H.-G.;
RA "Primary structure of rat RT6.2, a nonglycosylated
RT phosphatidylinositol-linked surface marker of postthymic T cells.";
RL PROC. NATL. ACAD. SCI. U.S.A. 87:964-967(1990).
[2]
RN
RP SEQUENCE OF 1-201 FROM N.A.
RC STRAIN-DA; TISSUE=SPLEEN;
RX MEDLINE; 96355006.
RA HAAG F., KUHLBAUMER G., KOCH-NOLTE F., WINGENDER E., THIELE H.-G.;
RT "Structure of the gene encoding the rat T cell ecto-ADP-
RT ribosyltransferase RT6.";
RL J. IMMUNOL. 157:2022-2030(1996).
[3]
RN PARTIAL SEQUENCE.
RX MEDLINE; 90215814.
RA KASHAN A., BUCK F., HAAG F., KOCH F., THIELE H.-G.;
RT "A single-step purification procedure and partial amino acid sequence
RT analysis of picomole amounts of the rat T cell alloantigen RT6.2";
RL IMMUNOL. LETT. 23:133-138(1989).
[4]
RN CHARACTERIZATION.
RX MEDLINE; 94193612.
RA TAKADA T., IIDA K., MOSS J.;
RT "Expression of NAD glycohydrolase activity by rat mammary
RT adenocarcinoma cells transformed with rat T cell alloantigen RT6.2.";
RJ. BIOL. CHEM. 269:9420-9423(1994).
CC -1- FUNCTION: HAS BOTH NAD+ GLYCOHYDROLASE AND ADP-RIBOSYLTRANSFERASE
CC ACTIVITY (TO A LESSER EXTENT)
CC -1- CATALYTIC ACTIVITY: NAD(+)+ L-ARGININE = NICOTINAMIDE +
CC N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -1- CATALYTIC ACTIVITY: NAD(+)+ H(2)O = NICOTINAMIDE + ADP-RIBOSE.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -1- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensel@isb-sib.ch).

DR EMBL; M85193; G205806; --
DR EMBL; X99123; E257753; --
DR EMBL; X99122; E257751; --
DR PIR; A34866; A34866.
DR PROSITE; PS01291; ART; 1.
DR PFAM; PF01129; ART; 1.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
KW T-CELL DIFFERENTIATION; GPI-ANCHOR.
FT SIGNAL 1 20
FT CHAIN 21 246
FT PROPEP 247 275
FT FT
FT LPID 246 246
FT ACT_SITE 209 209
FS SEQUENCE 275 AA; 31438 MW; DBEAB84E CRC32;

Query Match 2.4%; Score 102; DB 1; Length 275;
Best Local Similarity 30.6%; Pred.No. 1.01e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;

Db 194 LGVYIKFEFS-FRPDQEP-VLPVGVEYQKVQTGYNEIFLDSPKRKSNCLNYSSAGAR 251

QY 131 LGE-MRGYAPSPDENSESLVLFEGDEVYSTIRKQYNGK-IPFRFIRGESE-LYTSDTV 187
Db 252 ES 253
QY 188 QN 189

RESULT 6
ID NRT1-RAT STANDARD; PRT; 275 AA.
AC P17982;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 (T-CELL MONO(ADP-
NAD(P)(+))-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL SURFACE PROTEIN
RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1)
DE RT6.1).
GN RATTUS NORVEGICUS (RAT).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEWIS A;
RX MEDLINE; 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell
allcoantigen RT6.1.";
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
RN [2]
RP MUTAGENESIS OF GLN-207.
RX MEDLINE; 96275529.
RA MAEHAMA T., HOSHINO S.-I., KATADA T.;
RT "Increase in ADP-ribosyltransferase activity of rat T lymphocyte
allcoantigen RT6.1 by a single amino acid mutation.";
RL FEBS LETT. 388:189-191(1996).
CC -!- FUNCTION: HAS NAD+ GLYCOHYDROLASE ACTIVITY AND EXTREMELY LOW ADP-
RIBOSYLTRANSFERASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -!- CATALYTIC ACTIVITY: NAD(+) + H(2)O = NICOTINAMIDE + ADP-RIBOSE.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -!- PFM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X52082; G57168; -.
DR EMBL; M31138; G206804; -.
DR PIR; S08464; S08464.
DR PROSITE; PS01291; ART; 1.
DR PFM; PF01129; ART; 1.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
T-CELL DIFFERENTIATION; GPI-ANCHOR.
FT SIGNAL 1 20
FT CHAIN 21 246
FT PROPEP 247 275
FT T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
FT HYDROPHOBIC, REMOVED DURING MATURATION
FT (BY SIMILARITY).
FT GPI-ANCHOR (BY SIMILARITY).
FT LIPID 246 246
FT ACT SITE 209 209
FT CARBOHYD 58 58
FT POTENTIAL.
FT Q->E: INCREASED ADP-RIBOSYLTRANSFERASE
FT ACTIVITY.
SQ SEQUENCE 275 AA; 31388 MW; 523B1A84 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 2.90e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;
Db 194 LGVIKKEFS-FYPDQEE-VLIPGYEYQKVRTQGYNEIFLDSPPKRSKSNYNCLYSSAGTR 251
QY 131 LGE-MRGYAPSPDENSESLVLFEGDEVYSTIRKQYNGK-IPFRFIRGESE-LYTSDTV 187
Db 252 ES 253
QY 188 QN 189

RESULT 7
ID EPIA-TRE STANDARD; PRT; 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEF1.
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-QM9414 / RUTC-30;
RX MEDLINE; 94124021.
RA NAKARI T., ALATAO E., PENITIL M.;
RT "Isolation of Trichoderma resei genes highly expressed on glucose-
containing media: characterization of the tef1 gene encoding
translation elongation factor 1 alpha.";
RL GENE 136:313-318(1993).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; Z23012; G312887; -.
DR PIR; S35772; S35772.
DR PROSITE; PS00301; EFACTOR_GTP; 1.
DR PFM; PF00009; GTP_EFTU; 1.
DR HSP; P07157; IAP.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 949830 MW; B9ABAB2 CRC32;
Query Match 2.4%; Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 2.90e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;
Db 210 YGWEKETKAGKFTG-KTLLEAIDSIEPPKR-PTDKPLRLPLQDV 252
QY 478 YCGWDQGRCSISYSERSVLSQINPAEPKPCPNKPKAPLQKV 522

RESULT 8
ID VRP2-SALTY STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)


```
Best Local Similarity 23.3%; Pred. No. 2.90e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHTAQW-LVEESVTPAGE-HIYYSYLAENDNVN 219
QY 92 RDCENYITLLRERSEGLLACGTNARHPSCNWLVNGTVVPLGEMRGYAPFSPDENS 151
Db 220 GNEAGDRSARMYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPRRIRGESELY 181

RESULT 11
ID VRP2_SALDU STANDARD; PRT; 593 AA.
AC P24419;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN VSDC.
OS SALMONELLA DUBLIN.
OG PLASMID PSDL2.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC SALMONELLA.
CC [1]
CC SEQUENCE FROM N.A.
RR MEDLINE: 91251759.
RA KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINEY D.;
RT "Molecular analysis of the virulence locus of the Salmonella dublin
RT Plasmid PSDL2.";
RL MOL. MICROBIOL. 5:307-316(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC -!- OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X56727; G47839; .
DR PIR: S15215; S15215.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 375 POLY-PRO.
SQ SEQUENCE 593 AA; 65613 MW; A7CE57B5 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 593;
Best Local Similarity 23.3%; Pred. No. 2.90e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHTAQW-LVEESVTPAGE-HIYYSYLAENDNVN 219
QY 92 RDCENYITLLRERSEGLLACGTNARHPSCNWLVNGTVVPLGEMRGYAPFSPDENS 151
Db 220 GNEAGDRSARMYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPRRIRGESELY 181

RESULT 12
ID PNA1_AJCA STANDARD; PRT; 916 AA.
AC Q07421;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).
GN PNA1.
OS AJELLOWYCES CAPSULATA (HISTOPLASMA CAPSULATUM).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;

Best Local Similarity 23.3%; Pred. No. 2.90e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;

Db 162 HDSNGILHLGKTAARLSDPQAASHTAQW-LVEESVTPAGE-HIYYSYLAENDNVN 219
QY 92 RDCENYITLLRERSEGLLACGTNARHPSCNWLVNGTVVPLGEMRGYAPFSPDENS 151
Db 220 GNEAGDRSARMYLSKV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPRRIRGESELY 181

RESULT 13
ID HA21_HUMAN STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
```

```
ONYGENALES; ONYGENACEAE; AJELLOWYCES.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 94124018.
RA SCHAFER M.P., DEAN G.E.;
RT "Cloning and sequence analysis of an H(+)-ATPase-encoding gene from
RT the human dimorphic pathogen Histoplasma capsulatum.";
RL GENE 136:295-300(1993).
CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING
CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKINIZATION MAY MEDIATE
CC GROWTH RESPONSES.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L07305; G409249; .
DR PROSITE: PS00154; ATPASE_E1E2; 1.
DR PFAM: PF00122; E1-E2-ATPase; 1.
KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
KW ATP-BINDING.
FT DOMAIN 1 111 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 112 134 1 (POTENTIAL).
FT DOMAIN 135 136 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 137 155 2 (POTENTIAL).
FT DOMAIN 156 279 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 280 305 3 (POTENTIAL).
FT DOMAIN 306 320 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 321 350 4 (POTENTIAL).
FT DOMAIN 351 711 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 712 734 5 (POTENTIAL).
FT DOMAIN 735 754 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 755 776 6 (POTENTIAL).
FT DOMAIN 777 817 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 818 843 7 (POTENTIAL).
FT DOMAIN 844 847 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 848 874 8 (POTENTIAL).
FT DOMAIN 875 916 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 33 62 ASP/GLU-RICH (ACIDIC).
FT BINDING 374 374 PHOSPHORYLATION (BY SIMILARITY).
FT BINDING 470 470 ATP (BY SIMILARITY).
SQ SEQUENCE 916 AA; 98884 MW; 4A468A44 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 916;
Best Local Similarity 37.2%; Pred. No. 2.90e-01;
Matches 16; Conservative 8; Mismatches 16; Indels 3; Gaps 3;

Db 762 LAVGTWIT-LTTMLVSGENGIVQNFGRTHPVLFLFSLTENW 803
QY 244 LSVSKNTFLKAMLVCS-D-AATNKNFNRLQDV-FLLPDPGQW 284

RESULT 13
ID HA21_HUMAN STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
```


THIS PAGE BLANK (USPTO)

W P R E H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:56:43 1999; MasPar time 36.96 Seconds
828.471 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-236-2
Description: (41-601) from US09041236.pep (43 of 45)
Perfect Score: 4189
Sequence: 1 EPHTVLFHPGSSSVWVGGR.....BAQEGSYFRAQHWQLLPED 561

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertibrate 14:sp.virus

Statistics: Mean 48.540; Variance 75.079; scale 0.647

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					PRED. NO.	
Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	4189	100.0	666	4	075326 SEMAPHORIN L.	0.00e+00
2	2219	53.0	393	11	088371 SEMAPHORIN L. (FRAGMENT	0.00e+00
3	1825	43.6	653	14	064906 SIMILAR TO GENBANK ACC	0.00e+00
4	737	17.6	748	11	062177 SEMAPHORIN A PRECURSOR	5.26e-154
5	725	17.3	751	11	062181 SEMAPHORIN E PRECURSOR	7.48e-151
6	716	17.1	749	4	013214 SEMAPHORIN V.	1.72e-148
7	716	17.1	750	4	093018 SEMAPHORIN V.	1.72e-148
8	710	16.9	751	4	099985 SEMAPHORIN E.	6.46e-147
9	705	16.8	751	13	042236 COLLAPSIN 3.	1.32e-145
10	688	16.4	753	4	013372 SEMAPHORIN III FAMILY	3.74e-141
11	682	16.3	754	11	088633 SEMAPHORIN IV ISOFORM	1.39e-139
12	681	16.3	772	13	090607 COLLAPSIN.	2.53e-139
13	682	16.3	785	11	088632 SEMAPHORIN IV ISOFORM	1.39e-139
14	681	16.3	785	4	013275 SEMAPHORIN IV.	2.53e-139
15	679	16.2	785	4	015704 SEMAPHORIN.	8.45e-136
16	669	16.0	772	11	062180 SEMAPHORIN D PRECURSOR	3.46e-136
17	668	15.9	772	11	063548 SEMAPHORIN III/COLLAPS	6.31e-136
18	661	15.8	771	4	014563 SEMAPHORIN-III.	4.24e-134
19	618	14.8	782	11	062179 SEMAPHORIN C (SEM C)	6.54e-123
20	614	14.7	775	11	P70275 SEMAPHORIN H.	7.13e-122

21	599	14.3	861	11	009126 SEMAPHORIN J (SEMAPHOR	5.49e-118
22	590	14.1	761	13	090663 COLLAPSIN-2.	1.17e-115
23	588	14.0	775	4	015041 KIAA0331.	3.84e-115
24	572	13.7	785	13	042237 COLLAPSIN 5.	5.17e-111
25	568	13.6	294	13	090664 COLLAPSIN-3 (FRAGMENT	5.55e-110
26	544	13.0	862	4	092854 SEMAPHORIN.	8.22e-104
27	522	12.5	834	11	064151 SEMAPHORIN I (M-SEVA F	3.54e-98
28	511	12.2	299	13	090666 COLLAPSIN-5 (FRAGMENT	2.28e-95
29	480	11.5	760	11	062178 SEMAPHORIN B PRECURSOR	1.73e-87
30	473	11.3	295	13	090665 COLLAPSIN-4 (FRAGMENT	1.03e-85
31	447	10.7	730	5	026473 FASCICLIN IV.	3.70e-79
32	448	10.7	1074	4	013591 SEMAPHORIN F HOMOLOG.	2.07e-79
33	441	10.5	1077	11	062217 SEMAPHORIN F PRECURSOR	1.19e-77
34	415	9.9	1093	11	060519 SEMAPHORIN G PRECURSOR	2.13e-71
35	392	9.4	888	11	035464 SEMAPHORIN VIA.	1.93e-65
36	390	9.3	494	4	060408 SEMAPHORIN F (FRAGMENT	6.02e-65
37	378	9.0	712	5	026972 SEMAPHORIN-I PRECURSOR	5.48e-62
38	367	8.8	562	5	017330 CESEMA.	2.75e-59
39	367	8.8	771	5	024322 SEMAPHORIN-I.	2.75e-59
40	362	8.6	706	5	024323 SEMAPHORIN-II.	4.61e-58
41	341	8.1	284	11	054948 SEMAPHORIN-IV HOMOLOG	5.96e-53
42	316	7.5	886	11	054951 SEMAPHORIN N (SEMAPHOR	6.24e-47
43	316	7.5	887	11	070141 SEMAPHORIN Z.	6.24e-47
44	283	6.8	770	5	042553 SIMILAR TO SEMAPHORIN-	4.12e-39
45	204	4.9	1963	4	075051 KIAA0463 PROTEIN (FRAG	3.24e-21

ALIGNMENTS

RESULT	1				
ID	075326	PRELIMINARY;	PRT;	666	AA.
AC	075326;				
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	SEMAPHORIN L.				
GN	SEMA.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;				
OC	CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98389619.				
RA	LANG E C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;				
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA				
RT	viruses."				
RL	GENOMICS 51:340-350(1998).				
DR	EMBL; AF030698; G3523115; "				
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;				
Query Match	100.0%;	Score 4189;	DB 4;	Length 666;	
Best Local Similarity	100.0%;	Pred. No. 0.00e+00;			
Matches	561;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Db	73	EPHTVLFHPGSSSVWVGGRKYLFDPEFGKNASVTVNIGSTKSCLDKRCENYIL	132
QY	41	EPHTVLFHPGSSSVWVGGRKYLFDPEFGKNASVTVNIGSTKSCLDKRCENYIL	100
Db	133	LERRSGLLACGNNARHPSCWNLVNGTVVPLGEMRGYAFSPDENSLVLFEGDEVSTIR	192
QY	101	LERRSGLLACGNNARHPSCWNLVNGTVVPLGEMRGYAFSPDENSLVLFEGDEVSTIR	160
Db	193	KOEYNGKIPFRIRGESELYTSDTYMNPQFIKATIVHQDAYDDKIYFFREDNPKN	252
QY	161	KOEYNGKIPFRIRGESELYTSDTYMNPQFIKATIVHQDAYDDKIYFFREDNPKN	220
Db	253	PEAPLNVSRAQLCRDQGGESSLSVSKWNTFLKMLVCSDAATNKNRNLQDVFLLDP	312
QY	221	PEAPLNVSRAQLCRDQGGESSLSVSKWNTFLKMLVCSDAATNKNRNLQDVFLLDP	280
Db	313	SGQWRTRYGVGFSPNPNYSACVYSLGDDIKVFRFTSSLKGYHSSLNPNRPGCLDQDP	372
QY	281	SGQWRTRYGVGFSPNPNYSACVYSLGDDIKVFRFTSSLKGYHSSLNPNRPGCLDQDP	340

```
Db 373 IPTETFOVADRHPEVAQRVPEPMGLKTPFLFHSKYHYQKVAVHRMQASHGETFHVLYLTLD 432
Qy 341 IPTETFOVADRHPEVAQRVPEPMGLKTPFLFHSKYHYQKVAVHRMQASHGETFHVLYLTLD 400
Db 433 RGTIHKVVEPGQEHSAFNIMEIQPFRRAAIAQTMSLDAERKLYVSSQWEVSQVPLDL 492
Qy 401 RGTIHKVVEPGQEHSAFNIMEIQPFRRAAIAQTMSLDAERKLYVSSQWEVSQVPLDL 460
Db 493 CEVYGGCHGCLMSRDPYCGWQDQGCISYSSERSVLSQINPAEPHKECPNPKPKAPLQ 552
Qy 461 CEVYGGCHGCLMSRDPYCGWQDQGCISYSSERSVLSQINPAEPHKECPNPKPKAPLQ 520
Db 553 KVS LAPNSRYILSCPMESRHATYSWRHKNVQSCPEGHQSPNCILFIENLTAQYGHYF 612
Qy 521 KVS LAPNSRYILSCPMESRHATYSWRHKNVQSCPEGHQSPNCILFIENLTAQYGHYF 580
Db 613 CEAQEGSYFRAQHWOLLPED 633
Qy 581 CEAQEGSYFRAQHWOLLPED 601

RESULT 2
ID O88371 PRELIMINARY: PRT: 393 AA.
AC O88371:
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN L (FRAGMENT).
GN SEMAL.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT viruses."
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030699; G3523117; -.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43909 MW; 7C88193C CRC32;

Query Match 53.0%; Score 2219; DB 11; Length 393;
Best Local Similarity 92.6%; Pred. No. 0.00e+00;
Matches 300; Conservative 12; Mismatches 11; Indels 1; Gaps 1;

Db 70 EPHTVLFEPGSSVWVGGRGVYHFNPEPGKNASVRTVNIIGSTKGQDQKODCGNYITL 129
Qy 41 EPHTVLFEPGSSVWVGGRGVYHFNPEPGKNASVRTVNIIGSTKGQDQKODCGNYITL 100
Db 130 LERRGNLLVCTGNARKPSCNWLNVDSVVMISLGEKMGVAPFSPDENSLVLFEGDEVYSTI 189
Qy 101 LERRSEGLLACGTNARHPSCNWLNVGTVP-PLGEMRGVAPFSPDENSLVLFEGDEVYSTI 159
Db 190 RKOENYNGKIPRRFRIRGESELYTSDTMONPOFIKATIVHQDQAYDDKIIYFFREDNPK 249
Qy 160 RKOENYNGKIPRRFRIRGESELYTSDTMONPOFIKATIVHQDQAYDDKIIYFFREDNPK 219
Db 250 NPEAPLNVSRVAQLCRDQGGESSLSVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLLPD 309
Qy 220 NPEAPLNVSRVAQLCRDQGGESSLSVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLLPD 279
Db 310 PSGQRDRTRVYGVSFNPWNYSAVCYSLGIDRVFRTSSLKGYHGLNPRPGMCLPKQ 369
Qy 280 PSGQRDRTRVYGVSFNPWNYSAVCYSLGIDRVFRTSSLKGYHGLNPRPGMCLPKQ 339
Db 370 PIPTETFOVADRHPEVAQRVPEPMG 393
Qy 340 PIPTETFOVADRHPEVAQRVPEPMG 363
```

```
RESULT 3
ID Q64906 PRELIMINARY: PRT: 653 AA.
AC Q64906:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALCELAHPHINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-C500;
RX MEDLINE: 97201573.
RA ENSSER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-C500;
RX MEDLINE: 97404659.
RA ENSSER A., PELANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-C500;
RA ENSSER A., PELANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; U18243; G1000717; -.
DR EMBL; AF005370; G2337970; -.
SQ SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match 43.6%; Score 1825; DB 14; Length 653;
Best Local Similarity 47.5%; Pred. No. 0.00e+00;
Matches 262; Conservative 95; Mismatches 184; Indels 11; Gaps 11;

Db 95 EPHVTVLFHSLNSDDVYGGNNTIYLFDFAHSSNASTALINITSTHTNHRSLSTCENFTTL 154
Qy 41 EPHVTVLFHPEGSSVWVGGRGVYHFNPEPGKNASVRTVNIIGSTKGQDQKODCGNYITL 100
Db 155 LHNQTDGLLACGTNSQKPSCH-LINLNTTQFLGPKGLAPFSPGSGNLVLFQDNDYVSTI 213
Qy 101 LERSEGLLACGTNARHPSCNWLNVGTVP-LGEMRGVAPFSPDENSLVLFEGDEVYSTI 159
Db 214 NLYKSLGS-HKFRRIAGQVELYTSDTAMHRPQFQVATVHKNESYDDKIYFFEQENSHS 272
Qy 160 RK-OEYNGKIPRRFRIRGESELYTSDTMONPOFIKATIVHQDQAYDDKIIYFFREDNPD 218
Db 273 DFKQFPHTVPRVGVQCSDDQGGESSLSVYKWTTFELKARLACVDYDTGRIYNELQDIFWQ 332
Qy 219 KNPEAPLNVSRVAQLCRDQGGESSLSVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLLP 278
Db 333 APENSWEETLYIGLFLSPWNFSVAVCVFTVKDIDHVFKTSKLKYNHKLPTPRPGCKMNH 392
Qy 279 DPSQWRDRTRVYGVSFNPWNYSAVCYSLGIDRVFRTSSLKGYHGLNPRPGMCLPKDQ 338
Db 393 QHVTEITFOVADRHPEVAQRVPEPMGPIIOSKYITKLVYRVEYG-GVFWATFIYL 451
Qy 339 QPIPTETFOVADRHPEVAQRVPEPMGPIKPLTFHSHKYHYQKVAVHRMQASHGETH-VLYL 397
Db 452 TTIKGTHIYVREDSNSTALNILEINPFQKAPIQNLILDNTNLKLYVNSEVSEVP 511
Qy 398 TDRGTIHKVVEPGQEHSAFNIMEIQPFRRAAIAQTMSLDAERKLYVSSQWEVSQV 457
Db 512 LDLSVYNGDCFCFSMRDPLCTWYNNFC-S-FK-QRVSVETGGPANRTELSEMCDHYAP 568
Qy 458 LDLCVYGGCHGCLMSRDPYCGWQDQGCISYSSERSVLSQINPAEPH-KE-CPNPKPD 515
Db 569 TVVKHOYSIPLLSNLSYLSCLPAVSNHADYFWTKDGFTEKRCHVKTHKNDCILLIANSTAT 628
Qy 516 KAPLQKVS LAPNSRYILSCPMESRHATYSWRHKNVQSCPEGHQSPNCILFIENLTAQ 575
```



```

DT 01-MAY-1997 (TREMBREL. 03, CREATED)
DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA YAMADA T., ENDO R., GOTOH M., HIROHASHI S.;
RT "Identification of semaphorin E as a non-MDR drug resistance gene of
RT human cancers.";
RL PROC. NATL. ACAD. SCI. U.S.A. 94:14716-14718(1997).
DR EMBL: AB000220; D1033360; -.
SQ SEQUENCE 751 AA; 85207 MW; F89FB934 CRC32;

Query Match 16.9%; Score 710; DB 4; Length 751;
Best Local Similarity 33.6%; Pred. No. 6.46e-147;
Matches 151; Conservative 104; Mismatches 151; Indels 43; Gaps 30;

Db 167 SFNPNTVSVMINELFSGMYI-DFMGTDAAIFSLTKRNAVRTDOHNSKWLSEPFVD 225
QY PFPSPDENSELVLFEGDEVYSTIRKQYNGK-IPFRIRGESELYTS--DTV-MONPOFIK 194
[1]
Db 226 AHVIPDGTDPNDKAVFFKKEKLTDSNSTKQIHSMIARICPNDTGGLRSL-VNKKWTFEL 284
QY ATIVHOD-QAYDDKIYFFREDNPKNPEAPLNVSVAQLCRDQGESLSVSKWNTFL 253
[1]
Db 285 KARLVCSVDEDEGPEHFELEDFVLL-ETDNP-RITLVYGIPTTSSVFKGSACVYHL 342
QY KAMLVCS--DA-ATKNFNRLQDVFLLPDPSGQWRDTRVYGFSPWN-Y--SACVYSL 307
[1]
Db 343 SDIQTVFNGPFAHKEGPNHQLIPYQGRIPRPGTCGGAFTPMRTTKEFFDDVVFIR 402
QY GDIDKVERTS-SLK-G-----YHSSLNPRPGKLPDQ-QP-IPDET-F--QVADRH 351
[1]
Db 403 NHPLMNSIPIHTRKPLIVRI-GTDKYTKIAVDVNAADG-RYHVLFGTDGRTGTVKV 460
QY HPEVAQRVEPM-G-PLKTPLFHSKYHYQKAVHRMQASHGETFHVLYLTDRGTIHKV 408
[1]
Db 461 VLPTNNSVSGELILEEVEFKNHAPITTMKISSKKQOLYSSNEGVSVSLRCHRYGTA 520
QY E-PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERKLYSSQWESVQVPLDCEVYGGG 467
[1]
Db 521 CADCLLARDPYCAWDGSCSRFYPTGKRSSRDQVRHGNPLTQCRGNLKAIRNAAEIVQ 580
QY CHCGLMSRDPYCGWDQGRICISYSE--RSVLQINPAEPHKECP--NPKPDKAPLQKVS 523
[1]
Db 581 YGKNNITTELECAPKSPQASIKWLLQKDK 609
QY LAP-NSRYLSCPMSRHATYSWR-HKEN 550

RESULT 9
ID O42236 PRELIMINARY; PRT; 751 AA.
AC O42236;
DT 01-JAN-1998 (TREMBREL. 05, CREATED)
DT 01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)
DE COLLAPIN 3.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA FEINER L., KOPEL A.M., KOBAYASHI H., RAPER J.A.;
RL NEURON 19:0-0(1997).
DR EMBL: AF022946; G252204; -.
SQ SEQUENCE 751 AA; 85432 MW; 0B9BA2F7 CRC32;

Query Match 16.8%; Score 705; DB 13; Length 751;
Best Local Similarity 33.5%; Pred. No. 1.32e-145;
Matches 150; Conservative 107; Mismatches 150; Indels 41; Gaps 28;

Db 167 SFNPNTVSVMINELFSGMYI-DFMGTDAAIFSLTKRNAVRTDOHNSKWLSEPFVD 225
QY PFPSPDENSELVLFEGDEVYSTIRKQYNGK-IPFRIRGESELYTS--DTV-MONPOFIK 194
[1]
Db 226 AHVIPDGTDPNDKAVFFKKEKLTDSNSTKQIHSMIARICPNDTGGLRSL-VNKKWTFEL 284
QY ATIVHOD-QAYDDKIYFFREDNPKNPEAPLNVSVAQLCRDQGESLSVSKWNTFL 253
[1]
Db 285 KARLVCSVDEDEGPEHFELEDFVLL-ETDNP-RITLVYGIPTTSSVFKGSACVYHL 342
QY KAMLVCS--DA-ATKNFNRLQDVFLLPDPSGQWRDTRVYGFSPWN-Y--SACVYSL 307
[1]
Db 343 SDIQTVFNGPFAHKEGPNHQLIPYQGRIPRPGTCGGAFTPMRTTKEFFDDVVFIR 402
QY GDIDKVERTS-SLK-G-----YHSSLNPRPGKLPDQ-QP-IPDET-F--QVADRH 351
[1]
Db 403 NHPLMNSIPIHTRKPLIVRI-GTDKYTKIAVDVNAADG-RYHVLFGTDGRTGTVKV 460
QY HPEVAQRVEPM-G-PLKTPLFHSKYHYQKAVHRMQASHGETFHVLYLTDRGTIHKV 408
[1]
Db 461 VLPTNNSVSGELILEEVEFKNHAPITTMKISSKKQOLYSSNEGVSVSLRCHRYGTA 520
QY E-PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERKLYSSQWESVQVPLDCEVYGGG 467
[1]
Db 521 CADCLLARDPYCAWDGSCSRFYPTGKRSSRDQVRHGNPLTQCRGNLKAIRNAAEIVQ 580
QY CHCGLMSRDPYCGWDQGRICISYSE--RSVLQINPAEPHKECP--NPKPDKAPLQKVS 523
[1]
Db 581 YGKNNITTELECAPKSPQASIKWLLQKDK 609
QY LAP-NSRYLSCPMSRHATYSWR-HKEN 550

RESULT 10
ID O13372 PRELIMINARY; PRT; 753 AA.
AC O13372;
DT 01-NOV-1996 (TREMBREL. 01, CREATED)
DT 01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN III FAMILY HOMOLOG.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-BRAIN;
RX MEDLINE: 96230324.
RA XIANG R.H., HENSEL C.H., GARCIA D.K., CARLSON H.C., KOK K., DALY M.C.,
RA KEBACHER K., DEN BERG A., VELDHIJS P., BUYS C.H., NAYLOR S.L.;
RT "Isolation of the human semaphorin III/F gene (SEMA3F) at chromosome
RT 3p21, a region deleted in lung cancer.";
RL GENOMICS 32:39-48(1996).
DR EMBL: U38276; G1061351; -.
DR PFAM: PF00047; Ig; 1.
SQ SEQUENCE 753 AA; 84941 MW; BECBEBB0 CRC32;

Query Match 16.4%; Score 688; DB 4; Length 753;
Best Local Similarity 31.4%; Pred. No. 3.74e-141;
Matches 139; Conservative 121; Mismatches 140; Indels 42; Gaps 30;

Db 171 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFSLTKRNAVRTDOHNSKWLSEPFV 229
QY PFPSPDENSELVLFEGDEVYSTIRKQYNGK-IPFRIRGESELYTS--DTV-MONPOFIK 194
[1]
Db 230 AELIPDSANDDKLFFFRSAAE-APQSPAVYARIGRLDNDGHCCL-VNKKWTFELK 287
QY ATIVHODQAYDDKIYFFREDNPKNPEAPLNVSVAQLCRDQGESLSVSKWNTFLK 254
[1]
Db 288 ARLVCSVPGEDGIEHFELEDFVLL-ETDNP-RITLVYGIPTTSSVFKGSACVYSL 345
QY KAMLVCS--DA-ATKNFNRLQDVFLLPDPSGQWRDTRVYGFSPWN-WNY--SACVYSLG 308
[1]
```



```

RN  [1]
RP  SEQUENCE FROM N.A.
RA  ECKHARDT F., MEYERHANS A.;
RT  "Molecular cloning and expression pattern of a murine semaphorin
RL  homologous to H-sema IV.";
RL  SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL; AF080090; G3377766;
SQ  SEQUENCE 785 AA; 88493 MW; 6FFCD8F5 CRC32;

Query Match      16.3%; Score 682; DB 11; Length 785;
Best Local Similarity 31.6%; Pred. No. 1.39e-139;
Matches 141; Conservative 119; Mismatches 140; Indels 43; Gaps 31;

Db  202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRLGKOTAMRTDOYNSRWLNDPSFIH 260
QY  139 PFPDENSELVFEGDEVYTIKQYNGK-IPRFRIRGESELYTS--DTV-MQNPOFIK 194
Db  261 AELIPDSAEKNDKLYFFFRERSAE-APQNPAYVARIGRICLNDGGHCL-VNKNSTFL 318
QY  195 ATIVHOD-QAYDDKIYFFREDNPKNPEAPLNVSVAQLCRDQGGESSLSVKNWTF 253
Db  319 KARLVCSVPEGDGIETHFDELQDFV-QQTQDI-RNPVIYAVFTSSGVSFRGSVAVCYSM 376
QY  254 KAMLVCS--DAATNKNFNRLQDVLLPDPGQWRDTRVYGVFSPN-WNY--SAVCVYSL 307
Db  377 ADTRMVFNGPFAHKEGPNYQWMPFSGKMPYPRGTCGGTFTFSMKSTKDYDDEVINMR 436
QY  308 GDIDKVFRTS-SLK-G-YH----SS-LPNPRGKCLPDQ-QP-I-PTETF--QVAD--R 351
Db  437 THPLMYQAVYPLQRRPLVVRT-CAPYRLTTIYAVDOVDAADG-RYEVLFGLTDRGTQKVI 494
QY  352 -HPEVAQRVEPMG--PLKTPLFHSHKYHYQKVAVHRQASHGETFHVLYLTDRGTHKV 408
Db  495 VLPKDDQEEELMLEEVEFKPAPVKTMTISSKROQLYVASAVGVTLSLHRCQAYGAA 554
QY  409 E-PGEQHSFAFNIMEIQPFRRAAQTMSLDAERKLYVSSQWESVQVPLDICEVYGG 467
Db  555 CADCCCLARDPYCWDQACSRYSYASKRKRRODVHRGNPIROCRGFSNANKNAVESVQ 614
QY  468 CHGCLMSRDPYCGWDGRCIS-IYSSER-SVLQSIINPAEPHKECP--NPKPKAPLQKVS 523
Db  615 YGVAGSAAFLECPSPQATVKW 637
QY  524 LA-PNSRYLSCPMSRHATYSW 545

RESULT 14
ID  Q13275 PRELIMINARY; PRT; 785 AA.
AC  Q13275; Q13274;
DT  01-NOV-1996 (TREMBLREL. 01, CREATED)
DT  01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT  01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE  SEMAPHORIN IV.
OS  HOMO SAPIENS (HUMAN).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC  CATARRHINI; HOMINIDAE; HOMO.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE-BRAIN;
RX  MEDLINE; 96210603.
RA  SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA  ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT  "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT  cancer deletion region and demonstrate distinct expression
RT  patterns.";
RL  PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR  EMBL; AC000063; G1669373;
DR  EMBL; U32172; G995786;
DR  EMBL; U32171; G995786;

Query Match      16.2%; Score 679; DB 4; Length 785;
Best Local Similarity 31.6%; Pred. No. 8.45e-139;
Matches 140; Conservative 121; Mismatches 139; Indels 43; Gaps 31;

Db  202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRLGKOTAMRTDOYNSRWLNDPSFIH 260
QY  139 PFPDENSELVFEGDEVYTIKQYNGK-IPRFRIRGESELYTS--DTV-MQNPOFIK 194
Db  261 AELIPDSAEKNDKLYFFFRERSAE-APQNPAYVARIGRICLNDGGHCL-VNKNSTFL 318
```

```

DR  PFAM; PF00047; ig; 1.
SQ  SEQUENCE 785 AA; 88381 MW; CD175765 CRC32;

Query Match      16.3%; Score 681; DB 4; Length 785;
Best Local Similarity 31.4%; Pred. No. 2.53e-139;
Matches 139; Conservative 122; Mismatches 139; Indels 43; Gaps 31;

Db  202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRLGKOTAMRTDOYNSRWLNDPSFIH 260
QY  139 PFPDENSELVFEGDEVYTIKQYNGK-IPRFRIRGESELYTS--DTV-MQNPOFIK 194
Db  261 AELIPDSAEKNDKLYFFFRERSAE-APQNPAYVARIGRICLNDGGHCL-VNKNSTFL 318
QY  195 ATIVHOD-QAYDDKIYFFREDNPKNPEAPLNVSVAQLCRDQGGESSLSVKNWTF 253
Db  319 KARLVCSVPEGDGIETHFDELQDFV-QQTQDI-RNPVIYAVFTSSGVSFRGSVAVCYSM 376
QY  254 KAMLVCS--DAATNKNFNRLQDVLLPDPGQWRDTRVYGVFSPN-WNY--SAVCVYSL 307
Db  377 ADTRMVFNGPFAHKEGPNYQWMPFSGKMPYPRGTCGGTFTFSMKSTKDYDDEVINMR 436
QY  308 GDIDKVFRTS-SLK-G-YH----SS-LPNPRGKCLPDQ-QP-I-PTETF--QVAD--R 351
Db  437 SHPLMYQAVYPLQRRPLVVRT-CAPYRLTTIYAVDOVDAADG-RYEVLFGLTDRGTQKVI 494
QY  352 -HPEVAQRVEPMG--PLKTPLFHSHKYHYQKVAVHRQASHGETFHVLYLTDRGTHKV 408
Db  495 VLPKDDQEEELMLEEVEFKPAPVKTMTISSKROQLYVASAVGVTLSLHRCQAYGAA 554
QY  409 E-PGEQHSFAFNIMEIQPFRRAAQTMSLDAERKLYVSSQWESVQVPLDICEVYGG 467
Db  555 CADCCCLARDPYCWDQACSRYSYASKRKRRODVHRGNPIROCRGFSNANKNAVESVQ 614
QY  468 CHGCLMSRDPYCGWDGRCIS-IYSSER-SVLQSIINPAEPHKECP--NPKPKAPLQKVS 523
Db  615 YGVAGSAAFLECPSPQATVKW 637
QY  524 LA-PNSRYLSCPMSRHATYSW 545

RESULT 15
ID  Q15704 PRELIMINARY; PRT; 785 AA.
AC  Q15704;
DT  01-NOV-1996 (TREMBLREL. 01, CREATED)
DT  01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT  01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE  SEMAPHORIN.
OS  HOMO SAPIENS (HUMAN).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC  CATARRHINI; HOMINIDAE; HOMO.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE-BRAIN;
RX  MEDLINE; 96226360.
RA  ROCHE J., BOLDOG F., ROBINSON M., ROBINSON L., VARELLA-GARCIA L.,
RA  SWANTON M.H., WAGGONER B., FISHEL R., FRANKLIN W., GEMMILL R.,
RA  DRABKIN H.;
RT  "Distinct 3p21.3 deletions in lung cancer and identification of a new
RT  human semaphorin.";
RL  ONCOGENE 12:1289-1297(1996).
DR  EMBL; U33920; G1000207;
DR  DR PFAM; PF00047; ig; 1.
SQ  SEQUENCE 785 AA; 88385 MW; 9971F14E CRC32;

Query Match      16.2%; Score 679; DB 4; Length 785;
Best Local Similarity 31.6%; Pred. No. 8.45e-139;
Matches 140; Conservative 121; Mismatches 139; Indels 43; Gaps 31;

Db  202 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRLGKOTAMRTDOYNSRWLNDPSFIH 260
QY  139 PFPDENSELVFEGDEVYTIKQYNGK-IPRFRIRGESELYTS--DTV-MQNPOFIK 194
Db  261 AELIPDSAEKNDKLYFFFRERSAE-APQNPAYVARIGRICLNDGGHCL-VNKNSTFL 318
```

QY 195 ATIVHQD-QAYDDKIYFFEDNPNKNEAPLNVSRVAQLCRDQDGESSLSVSKWNTFL 253
Db 319 KARLVCSYPGEDGIEHDELODFV-QOTQDV-RNPVIYAVETSSGVSVERGSVCVYSM 376
QY 254 KAWLVCS---DAATNKNFRLQDVFLLPDPSCQWRDTRVYGFVSNP-WNY--SAVCVYSL 307
Db 377 ADIRMFNGPFAHKEGPNYQWMPFSGKMPYPYRPGTCPPGCTFTPSMKSTKDYDDEVINEMR 436
QY 308 GDIDKVERTS-SLK-G-YH---SS-LPNRPGKCLPDQ-QP-I-PTETF--QVAD--R 351
Db 437 SHPLMQAVYPLQRRPLVVRT-GAPYRLTTIAVQVQVADGQ-RYEVLFLGTDRTGTQKVI 494
QY 352 -HPEVAQRVEPMG--PLKTPLFHSKYHYQKVAVHRMQASHGETFHVLYLTTRGTIHKV 408
Db 495 VLPKDDQEMEELMLEVEVEFKDPAPVKTMTISSKRQOLYVASAVGVTHLSLHRCOAYGAA 554
QY 409 E-PGQEHSAFNIWEIOFFRAAAIQTMSLDAERRKLYVSSQWEVSQVPLDLCEVYGGG 467
Db 555 CADCCCLARDPYCAWDQACSRYTASSKRRSRRQDYVRHGNPIRQCRGFNSNANKNAVESVQ 614
QY 468 CHGCLMSRDPYCGWDGRCIS-IYSER-SVLOSINPAEPHKECP--NPKPDKAPLQKVS 523
Db 615 YGVAGSAAPFLECQPRSPQATVKW 637
QY 524 LA-PNSRYLYLSCPMESRHATYSW 545

Search completed: Thu Jul 8 19:57:26 1999
Job time : 43 secs.

WAPSELEH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 20:02:13 1999; MasPar time 27.33 Seconds
Tabular output not generated. 432.711 Million cell updates/sec

Title: >US-09-041-236-2
Description: (6-561) from US09041236.pap (44 of 45)
Perfect Score: 4130
Sequence: 1 WAAASAOQHLSRPRIFAV.....TYSNRHKNVEQCEPHQS 556
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 2126608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.085; Variance 147.483; scale 0.245

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Pred. No.
1	656	15.9	771	13	2.33e-52
2	615	14.9	477	13	3.20e-48
3	608	14.7	775	33	1.82e-47
4	606	14.7	861	22	2.58e-47
5	599	14.5	861	32	1.31e-46
6	538	13.0	862	22	1.72e-40
7	518	12.5	776	32	1.71e-38
8	463	11.2	441	13	5.01e-33
9	459	11.1	730	13	1.25e-32
10	428	10.4	587	32	1.43e-32
11	390	9.4	974	33	7.57e-26
12	378	9.2	712	13	1.12e-24
13	367	8.9	650	13	1.31e-23
14	362	8.8	724	13	4.02e-23
15	352	8.5	930	32	3.74e-22
16	325	7.9	888	25	1.49e-19

17	316	7.7	887	25	W19856	Rat semaphorin 2.	1.09e-18
18	296	7.2	929	32	W57259	Rat semaphorin Y.	8.75e-17
19	157	3.8	122	13	R71385	Varicella major virus S	4.01e-04
20	99	2.4	591	4	R23006	Protein transcribed f	1.82e+01
21	98	2.4	832	33	W61092	Taq DNA polymerase I	2.16e+01
22	101	2.4	2329	25	W25038	Partial BRCA2 cancer	1.30e+01
23	96	2.3	832	33	W61090	Taq DNA polymerase I	3.03e+01
24	93	2.3	4572	30	W52845	A. mediterranei rifam	4.99e+01
25	89	2.2	143	25	W25052	BRCA2 cancer suscepti	9.60e+01
26	91	2.2	478	1	R04881	Recombinant elastase.	6.93e+01
27	89	2.2	534	25	W25031	Partial BRCA2 cancer	9.60e+01
28	91	2.2	598	9	R48631	Sequence of nuclear r	6.92e+01
29	90	2.2	638	5	R03924	E. coli HSP (dnak).	8.16e+01
30	89	2.2	680	4	R23143	Mutant thermostable D	9.60e+01
31	89	2.2	757	4	R23141	Mutant thermostable D	9.60e+01
32	89	2.2	787	4	R23142	Mutant thermostable D	9.60e+01
33	92	2.2	832	33	W61091	Taq DNA polymerase I	5.88e+01
34	89	2.2	832	13	R76691	DNA-polymerase REM-T3	9.60e+01
35	89	2.2	832	12	R64272	T. aquaticus DNA-poly	9.60e+01
36	89	2.2	832	1	P90536	Purified native therm	9.60e+01
37	89	2.2	832	33	W61089	Taq DNA polymerase I	9.60e+01
38	89	2.2	832	13	R76690	Taq DNA-polymerase RE	9.60e+01
39	89	2.2	833	27	W24211	Cleavase DN nuclease.	9.60e+01
40	89	2.2	833	27	W24212	Cleavase DA nuclease.	9.60e+01
41	89	2.2	833	27	W24213	Cleavase DV nuclease.	9.60e+01
42	92	2.2	854	30	W56309	Cla II S-receptor ki	5.88e+01
43	92	2.2	855	6	R29815	S receptor kinase pro	5.88e+01
44	89	2.2	3418	26	W19211	Human breast cancer s	9.60e+01
45	89	2.2	3418	26	W23287	Human breast and ovar	9.60e+01

ALIGNMENTS

RESULT 1

ID R71380 standard; Protein; 771 AA.
AC R71380;
DT 21-NOV-1995 (first entry)
DE Human semaphorin III protein.
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW varicella major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN WO9507706-A.
PD 23-MAR-1995.
PF 13-SEP-1994; U10151.
PR 13-SEP-1993; US-121713.
PA (REGC) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI; 95-131177/17.
DR N-PSDB; Q87442.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2; page 60-63; 101pp: English.
CC The sequence of the human semaphorin III protein. The proteins
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87445) or varicella major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor binding
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA;

Query Match 15.9%; Score 656; DB 13; Length 771;

Best Local Similarity 31.2%; Pred. No. 2.33e-52;

Matches 143; Conservative 115; Mismatches 155; Indels 45; Gaps 35;

Db 166 rgkspydpklltasllldgelysgta-adfmgdrdfairtlghhhpirtedqhdrwindp 224

Db 462 kvitinygetewmeevileelqikfdapiismeisskrqlyigsasavqvrffhcdm 521
 Qy 406 KVPFGEQEHSEFARN-IME-IQPFRRRAAIIOTMSLDAERRLIYVSSQWSEVSQVPLDLCEV 463
 Db 522 ygsacacclardypcawdgiscsryypt 550
 Qy 464 YGGCHGCLMSRDPVCGWDQGRCSISYSS 492

RESULT

ID W17658 standard; Protein; 861 AA.
 AC W17658;
 DT 24-JUL-1997 (first entry)
 DE Mouse CD100 antigen.
 KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
 OS Mus sp.
 FH Key
 FT peptide
 FT label= Sig_peptide
 FT protein
 FT label= Mat_protein
 FT domain
 FT label= Semaphorin_domain
 FT domain
 FT label= Ig-like_domain
 FT domain
 FT label= Stalk_domain
 FT domain
 FT label= Transmembrane_domain
 FT domain
 FT label= Cytoplasmic_domain
 FT modified_site
 FT label= Phosphorylation
 FT /note= "putative tyrosine phosphorylation site"
 PN W09717368-A1.
 PD 15-MAY-1997.
 PF 12-NOV-1996; U18645.
 PR 09-NOV-1995; US556422.
 PA (DAND) DANA FARBER CANCER INST.
 PI Boussiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
 DR WPI: 97-280982/25.
 DR N-PSDB; T60666.
 PT Nucleic acid molecule encoding Cdl00 antigen - which stimulates
 PT leukocyte response, e.g. B cell aggregation, differentiation,
 PT survival and T cell proliferation
 PS Example 8; Page 86-89; 135pp; English.
 CC Mouse CD100 antigen (W17657) is a novel leukocyte semaphorin-like
 CC protein that stimulates a leukocyte response, including B cell
 CC aggregation, B cell differentiation, B cell survival and/or T cell
 CC proliferation. Its amino acid sequence was deduced from a cDNA
 CC clone (T60666) isolated from murine T cells. Human CD100 antigen
 CC (W17657) has also been identified. CD100 polypeptides and fusion
 CC proteins, nucleic acids, and host cells expressing Cdl00 can be
 CC utilised in diagnostic and therapeutic methods involving modulation
 CC of B and T cell responses, neuron axonal growth and immune cell-
 CC nerve cell interaction.
 SQ Sequence 861 AA;

Query Match 14.7%; Score 606; DB 22; Length 861;
 Best Local Similarity 32.4%; Pred. No. 2.58e-47;
 Matches 143; Conservative 105; Mismatches 145; Indels 47; Gaps 31;

Db 107 eclnyrvlqlpsslsyvcgtnafqptcdhnltsfklgksedgkrcpfdpahsyts 166
 Qy 93 DCENYITLLER-RSEGLLACGTNARHPSC--WNLVNGTVVPLGEM-RGYAFSPDENSLV 148
 Db 167 vmvggelysgts-yfnlgseplisrnnshsplrteyaipwlnepsfvadvikspdgpe 225
 Qy 149 LFEDEVYSTIRKQEYNGKIPFRFRIRGESELYTSDTV--MQNPQFIKATIVHQ--D--Q 202
 Db 226 geddkvyffftevseyefvfkmlprvarvckdgqglrtliq-kkwtstfkarlicskp 284

Qy 203 AYDDKIYFFREDNPKNEAPLNVSRAQLCRDQGGESLSVSKWNTFLKMLVCSDA 262
 Db 285 dsglvfnllqdvflrap-gl-kepvfyavftqplnnvglsavcaytlatveavfsrgky 342
 Qy 263 ATKNFNRLQDVFLLPDPGQWRDTRVYGVFSPWNY---SACVYSLGDDIKVF-R--- 315
 Db 343 mqsatveqstktwryngpvtprpgacidaeaaaytslnlnpdktlqfvkdhplmdd 402
 Qy 316 --TSSL-----K-G-YHSSSLPNRPGKCL-PDQOPIP-TETFOVADRHP-E-V-AQR-VE- 360
 Db 403 svtpidnrpklkkdvnytgivvdrtdqalqdfydvnmfistdrgalhkaviltkev--v 460
 Qy 361 PMGPLKT-P-LFHSKYHYQKVAVHRMQASHGETFHVLYLTTRDGTIIHKVVEPGEHSEFA 418
 Db 461 --ieetqlfdsepyvltllsskkrkrfvagsngvvqaplaifcekhgs-cedcvlard 517
 Qy 419 FNIMEIQPFRRRAAIIOTMSLDAER-RKL-YVSSQWSEVSQVPLDLCEVYGGCHGCLMSRD 476
 Db 518 pycawspaikacvtlhqeeas 538
 Qy 477 PYCGWDQG-R-CISYSSERS 495

RESULT

ID W58540 standard; Protein; 861 AA.
 AC W58540;
 DT 02-SEP-1998 (first entry)
 DE Human semaphorin.
 KW Human; semaphorin; diagnosis; nervous disease; immune disease;
 KW nerve extension inhibiting activity.
 OS Homo sapiens.
 PN J10155490-A.
 PD 16-JUN-1998.
 PF 27-NOV-1996; 332900.
 PR 27-NOV-1996; JP-332900.
 PA (SUMU) SUMITOMO SEIYAKU KK.
 DR WPI: 98-391044/34.
 DR N-PSDB; V31121.
 DR New human semafolin gene - useful in the diagnosis of nervous system
 DR PT and immune disorders
 PS Claim 1; Page 10-12; 15pp; Japanese.
 CC The present sequence represents human semaphorin (translated from the
 CC Japanese specification as semafolin). Semaphorin has nerve extension
 CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
 CC treatment and researches on nervous diseases and immune diseases.
 SQ Sequence 861 AA;

Query Match 14.5%; Score 599; DB 32; Length 861;
 Best Local Similarity 32.4%; Pred. No. 1.31e-46;

Matches 143; Conservative 105; Mismatches 146; Indels 47; Gaps 31;

Db 107 eclnyrvlqlpsslsyvcgtnafqptcdhnltsfklgksedgkrcpfdpahsyts 166
 Qy 93 DCENYITLLER-RSEGLLACGTNARHPSC--WNLVNGTVVPLGEM-RGYAFSPDENSLV 148
 Db 167 vmvggelysgts-yfnlgseplisrnnshsplrteyaipwlnepsfvadvikspdgpe 225
 Qy 149 LFEDEVYSTIRKQEYNGKIPFRFRIRGESELYTSDTV--MQNPQFIKATIVHQ--D--Q 202
 Db 226 geddkvyffftevseyefvfkmlprvarvckdgqglrtliq-kkwtstfkarlicskp 284
 Qy 203 AYDDKIYFFREDNPKNEAPLNVSRAQLCRDQGGESLSVSKWNTFLKMLVCSDA 262
 Db 285 dsglvfnllqdvflrap-gl-kepvfyavftqplnnvglsavcaytlatveavfsrgky 342
 Qy 263 ATKNFNRLQDVFLLPDPGQWRDTRVYGVFSPWNY---SACVYSLGDDIKVF-R--- 315
 Db 343 mqsatveqstktwryngpvtprpgacidaeaaaytslnlnpdktlqfvkdhplmdd 402
 Qy 316 --TSSL-----K-G-YHSSSLPNRPGKCL-PDQOPIP-TETFOVADRHP-E-V-AQR-VE- 360
 Db 403 svtpidnrpklkkdvnytgivvdrtdqalqdfydvnmfistdrgalhkaviltkev--v 460

QY	361	PMGLPT-P-LFHSKTHYQKVAHVRMQASHGETFHVLYLTDTTGRGTHKVVVEPGEHSPA	418			
Db	461	--icetqlfdfevptllllsskkgkfkvyagsgnvvgdapiafackhgs-cedcvlard	517			
QY	419	FNIMEIOPFRRAAIQMSLDAER-RKL-YVSSOWEVSQVPLDCEVYGGCGHCLMSRD	476			
Db	518	pycawspalkacvtlhqeeas	538			
QY	477	PYCGWDQG-R-CISYSSERS	495			
RESULT 6						
ID	W17657 standard; Protein; 862 AA.					
AC	W17657;					
DT	24-JUL-1997 (first entry)					
DE	Human CD100 antigen.					
KW	CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte; vaccine.					
OS	Homo sapiens.					
FH	Key					
FT	peptide					
FT	Location/Qualifiers					
FT	1..41					
FT	/label=" Sig_peptide					
FT	42..862					
FT	/label=" Mat_protein					
FT	42..553					
FT	/label=" Semaphorin_domain					
FT	554..630					
FT	/label=" Ig-like_domain					
FT	631..733					
FT	/label=" Stalk_domain					
FT	735..752					
FT	/label=" Transmembrane_domain					
FT	753..862					
FT	/label=" Cytoplasmic_domain					
FT	808..815					
FT	modified_site					
FT	/label=" Phosphorylation					
FT	/note=" putative tyrosine phosphorylation site"					
PN	W09717368-A1.					
PD	15-MAY-1997.					
PF	12-NOV-1996; U18645.					
PR	09-NOV-1995; U5-556422.					
PA	(DAND) DANA FARBER CANCER INST.					
PI	Bousiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;					
PI	WPI; 97-280982/25.					
DR	N-PSDB: T06665.					
PT	Nucleic acid molecule encoding CD100 antigen - which stimulates leukocyte response, e.g. B cell aggregation, differentiation, survival and T cell proliferation					
PT	Claim 7; Page 70-72; 135pp; English.					
PS	Human CD100 antigen (W17657) is a novel leukocyte semaphorin-like protein that stimulates a leukocyte response, including B cell aggregation, B cell differentiation, B cell survival and/or T cell proliferation. Its amino acid sequence was deduced from a cDNA clone (T60665) isolated by COS cell expression cloning. CD100 polypeptides and fusion proteins, nucleic acids, and host cells expressing CD100 can be utilised in diagnostic and therapeutic methods involving modulation of B and T cell responses, neuron axonal growth and immune cell-nerve cell interaction.					
SQ	Sequence 862 AA;					
Query Match 13.0%; Score 538; DB 22; Length 862;						
Best Local Similarity 29.7%; Pred No. 1.72e-40;						
Matches 148; Conservative 124; Mismatches 175; Indels 52; Gaps 3						
Db	107	ecinyirvlqplsatyvcgtnafgpcadhlntsfkflgknedkgrcpfdpahsyts	166			
QY	93	DCENYITLLRRSE-GLLAGCTNARHPSCWNLVNGTVVPLG--EM-RGYAPFSPDEN-SL	147			
Db	167	vmvdg-elysgts-ynflgsepilnsnshsplrtayaipwnepsvfvadwlrkspdp	224			
QY	148	VLPEGEVSTIRKQYNGKIPRRIRGESELYTSDTV--MONPOFIKATVHQ--D--	201			

QY 194 KATIVHODQAYDDKIYFFREDNPD-KNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNTF 252
Db 262 lksrlncsvpgdyfpyfneiqgtsdilegnyggq-vekllygvfttpvnsiggsavcafs 320
QY 253 LKAWLVCSDAATNK-NFNRLQDVF-LLPDP-SGQWRDTRVYGVFSNPWN-Y--SAVCVYS 306
Db 321 mksllesfdgpfkeqetmnsnwlavpslkvppeprpgcgvndstltpdsvsnfvkshlmd 380
QY 307 L-G--D-IDKVFRT--SSLKG-YHS--SL--PNPRPGKCLPDQOPIPTETFOVADRHPEVA 356
Db 381 eavpafft-zpilirislqyrfkdvqdrtpdgdqaydvlfigtddgkvikalnsasf 439
QY 357 QRVEPMGLTPLFH-S-KYHOKVAV-HRQASHGETFHVLYLTTRDGTIHKVWEPG-- 411
Db 440 dssdtvdsvveelqlvlpypvkvnlvymdgdsklvvvsdellaiklhrcgskit 499
QY 412 EOEHSFAFNIME-IOPFRRAAQTOM--SLDAERKLYVSSQWEVSOVPLDLCEVYG-G 466
Db 500 ncrecvslqdpycawdvnclktavgsdpws 530
QY 467 GCHGCLMSRDPYCGWDO-G-RCISIVSSERS 495

RESULT 10

ID W51314 standard; Protein: 587 AA.
AC W51314;
DT 08-SEP-1998 (first entry)
DE Human semaphorin W.
KW Human; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW Immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PN WO9815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; JP203549.
PA (SUMO) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-PSDB: V07281, V07282.
PT Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 90pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 10.48; Score 428; DB 32; Length 587;
Best Local Similarity 28.99; Pred. No. 1.43e-29;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgddeliffettsrafsyerikvrvarvcagdlgrktlq-qrwtflikadllcp 115
QY 201 DQAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKMLVCS 260
Db 116 gpehgrassvldqvairpelga-gtpifygfssqwegatisavcafrpqdirtvlnpg 174
QY 261 DAATNKNFNRLQDVLPLDPSPGQWRDTRVYGVFSNPWN--YSAVCVYSIGDIDKVFRTS 317
Db 175 frelkhdcnrglpvndvdpqprgcceitnmklrhfgsslsipdrvtlfrhplmdrp 234
QY 318 --SLK-GYHSLP--N----PRPGKCLPDQOPIP--TETFOVADRHPE-V-AQR-VE-P 361
Db 235 vfpadghplltvtdtavlrvavrvtslsqkeydvlylgtedghlhravriga-ls-vl 292
QY 362 MGPLKT-PLFHSK-YHOKVAVHRMQASHGETFHVLYLTTRDGTIHKVWEPGQEHSAF 419
Db 293 edial--fpepgvnmkl-yhsw-llvgsrtvgtvnttcnrglqs-csecialadpvc 347

QY 420 NIMEIQFFRAAAIQTMSLDAERRKLYVSSQWEVSOVPLDLCEVYGGGCHGLMSRDPYC 479
Db 348 avsfridecvahagehrglvqdiesadvslcpkegerpvvfevqvataahvllpcsp 407
QY 480 GWD-Q-GRCISIVSSERSVLQSNPAEPHKECPNPKAPDKOVLKSLAPNSRYILSCPME 537
Db 408 sawascvwhqpsgv 421
QY 538 SRHATYSWRHKNV 551
RESULT 11
ID W64221 standard; Protein: 974 AA.
AC W64221;
DT 06-OCT-1998 (first entry)
DE Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN WO9827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GEMT) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-362774/31.
DR N-PSDB: V44295.
PT New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17j; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 9.4%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 7.57e-26;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechmfikvlkknddalfvogtnafnpscrnykndtlepfdfgsmarcpydakhn 172
QY 91 KRDCENYIT-LLERRSEGLLACGTNARHPSCWNLVNGTVVPLG-EMRGYA--PFSPDENS 146
Db 173 valfadgklysatvt-dflaidaviysl-geep--tlrtvkhdskwlkopyfvqadvy 228
QY 147 LVLFEGEVEYITRKQENYNG-KIPFRRIRGESELYTSDTMQONFOFIKATIVHQOAYD 205
Db 229 dylyffirelaveynlmgkvvprvaqckndmgsgsrqlkqwtflkarlncsyvpgds 288
QY 206 DKTYIFFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKAWLVCSDAATN 265
Db 289 hfyfnllqavtdvrlng--rdv-vlatfstpynslpgsavcaydmlldiasvftgrkeq 345
QY 266 K-NFNRLQDVLPLDPSPGQWRDTRVYGVFSNPWN-Y--SAVCVYSLGDDIKVF--RTSSL 319
Db 346 kspdstwtvypdervpdkprpgccagsssleryatsnefdddtlnfiktphlmdaevpsif 405
QY 320 KGYHSS--LPNPRPGKCLPD--QQPIPTETFOVADRHPEVA-Q--RVEP-M--G-P-L- 365

Oy 471 CLMSRDPYCGWDQGR-CISIYSSE-RSVLQINPAE-PHKECPN 511

Search completed: Thu Jul 8 20:02:47 1999
Job time : 34 secs.

THIS PAGE BLANK (USPTO)

Matches	139;	Conservative	121;	Mismatches	140;	Indels	42;	Gaps	30
Db	171	YPDPKLDTASALINEELYAGYII-DFMGTDAAIFRTLGKQTAMRTDOYNRWLNDPSFIH	229						
Qy	139	PFSDENSLFLFEGDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNPOFIK	194						
Db	230	AEIIPDSAENDDKLYFFERSAB-APQSPAVYARIGRLCNDGDDGGHCL-VNKKSTFLK	287						
Qy	195	ATIVHODQAYDDKIYFFEDNPNKPEAPLVNSRVQAQLCRGDQGGESSLSVKWNTFLK	254						
Db	288	ARLVCSVPGEDGTETHELDQDFV-OOTQDV-RNPYIYAVFTSSSGSVFSGVACVYSMA	345						
Qy	255	AMLVCS--DAATNKNFNRLQDVFLLPDPGQWRDTRVGVFSNP-WNY--SAVCVYSLG	308						
Db	346	DIRMVFNPFAPKEGPNYOMPFSGKMPYRPGTCPGGTFTPSMKKSTKDYDVEINFMRS	405						
Qy	309	DIDKVFRTS-SLK-G-YH---SS-LPNRPGKCLPDQ-QP-I-PTETF-QVAD-R-R-	351						
Db	406	HPLMYQAVYPLQRPLVVRT-GAPYRLTTTAVDOVDSADG-RVEVLGLDTRGTQKVIV	463						
Qy	352	HPEVAQRVEPMG--PLKTPFLFHSKYHYQKVAVHRMQASHGETFHLVLTTRDGTIHRVVE	409						
Db	464	LPKDDQEMEELMLEEVEVEFKDPAVPKVTMTSSKRQOOLYVASAGVTHLSLHRCQAYGAAC	523						
Qy	410	-PGEQHSFAFNTNEIQPFRAAAIQMTSDAERKLYVSSQEWESOVPLDLCEVYGGC	468						
Db	524	ADCLARDPYCAMDQACSRYTASSKRRSRQDVRHGNPIRQCQFGNSNANKNAVESVQY	583						
Qy	469	HGCLMSRDPYCGWDGRCIS-IYSSER-SVLQISINPAEPHKECP--NPKDPKAPLOKVS	524						
Db	584	GVAGSAAFLECPQSPQATVKW	605						
Qy	525	A-PNSRYILSCPMESRHATYSW	545						
RESULT	6								
ENTRY	A49069	#type complete							
TITLE	collapsin - chicken								
ORGANISM	#formal_name Gallus gallus #common_name chicken								
DATE	07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 04-Sep-1998								
ACCESSIONS	A49069								
REFERENCE	A49069								
#authors	Luo, Y.; Raible, D.; Raper, J.A.								
#journal	Cell (1993); 75:217-227								
#title	Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones.								
#accession	A49069								
#status	preliminary; not compared with conceptual translation								
#molecule_type	mRNA								
#residues	1-772	#label LUO							
#crossover-references	GB:U02528; NID:g410078; PID:g410079								
CLASSIFICATION	#superfamily semaphorin								
SUMMARY	#length 772 #molecular-weight 88867 #checksum 9712								
Query Match	16.5%;	Score 681;	DB 2;	Length 772;					
Best Local Similarity	31.5%;	Pred. No. 4.49e-121;							
Matches	146;	Conservative 119;	Mismatches 150;	Indels 48;	Gaps 38				
Db	166	RGKSPYDPKLLTASLLVDGELYSTA-ADPMGRDFAIFRLGHHHPITRQHDHSLNDP	224						
Qy	135	RGYAPFSDENSLVLFEGDEVYSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNPN	190						
Db	225	RFISAHILIPESDNPDDDKIVFFPRENAIDGHTGKATHARIGQICKNDFGGHRSL-VNKK	283						
Qy	191	QFIKATVHQ-DQAYDDKIYFFREDPNKPEAPLVNSRVQAQLCRGDQGGESSLSVKW	249						
Db	284	TTFLKARLICSVPGNGIDTHFDELQVFLM-NSKDP- KNP IVYGVFTTSSNIFKGSVC	341						
Qy	250	NTEFLKAMLVCS -DA-ATNKNFNRLQDVFLLPDPGQWRDTRVGVFSNPWN-Y-SAVC	303						
Db	342	MYSTWDRVRFLGYAHRDGNVQWVYQGRVYPRPGTCPSKTFFGGFSDTKDLPDEVIT	401						

```
Qy 304 VYSLGIDIKVFRTS-SLK-G--YH--S--S-LPNRPRGKCLPDQ-QPI-PTETF--QV-A 349
Db 402 FARSHPAMYNVPFINS-RPIMIKTDVDYQFTQIVVDRVDAEDGQ-YDVMFICTDIGTIVL 459
Qy 350 -DR-HPEVAQRVEPMGPKLTPLFHSK--YHYQKVAVHRMQASHGETFHVLYLTDRGTIH 405
Db 460 KVVSTPKETWHELEEVLLSEMVERPTVISAMKISTKQOQLYIGSATGVSOPLHRCDV 519
Qy 406 KVEE-PGEQHSFAFIME-IQFRAAAIQTMSLDAERRKLYSSQWESVQVPLDLCEV 463
Db 520 YKACAECLARDPYCAWDGSSCSRYPFTAKRTRRQDIRNGDPLTHCSDLQHDHNPSCQ 579
Qy 464 YGGGCHGCLMSRDPYCGWDGRCISYSS-RSVL-QSINPAEPHKECPN-PKPDKAP-- 518
Db 580 TLEEKIYGVENSSTFLECSQRSRAIVY-WQFQKONDDHKVE 621
Qy 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQSC 556

RESULT 7
ENTRY I48747 #type complete
TITLE semaphorin D - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I48747
REFERENCE #authors Puschel, A.W.; Adams, R.H.; Betz, H.
#journal Neuron (1995) 14:941-948
#title Murine semaphorin D/collapsin is a member of a diverse gene
family and creates domains inhibitory for axonal extension.
#cross-references MUID:95267431
#accession I48747
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type mRNA
#residues 1-772 #label RES
#cross-references EMBL:X85993; NID:g854329; PID:g854330
GENETICS semD
#gene #superfamily semaphorin
CLASSIFICATION #length 772 #molecular-weight 88710 #checksum 1776
SUMMARY

Query Match 16.2%; Score 669; DB 2; Length 772;
Best Local Similarity 30.2%; Pred.No. 2.42e-118;
Matches 170; Conservative 143; Mismatches 182; Indels 67; Gaps 51;

Db 58 HTFLDE-ERSRLYVGAKDHIFSFLNLIKDKQIWPVSY-TRRDECKWAGKOILKECA 115
Qy 43 HTVLPEPGSSVMYGGKVKYLFDFPEGKN-AS-VRTNIGSTKGC--LDK---RDCE 95
Db 116 NFIKLEAYNOTHLVACGTFPHCTIEVGHHPEDNIFKLQDSHFENGKSPYDPKL 175
Qy 96 NITLLRSEGL-LIACGTNARHPCWNL-V--N-GTVVPL--G--EM-RGAYAFSPDE 144
Db 176 LFTASLLIDGELYSGTA-ADFMRGDAIFRTLGDHHPITRTEQDHSRLNDPRFISAHLIPE 234
Qy 145 NSLVLFEGDEVYSTIRKQYNGK-IPRERRIRGESELYTS--DIV-MQNPQIKATIVHQ 200
Db 235 SNPPEDKVIYFFERNATGGEHSGKATHARIQICKNDFGGHRSI-VNKWTTFLKARLIC 293
Qy 201 -DQAYDDKIYFFERNDPKNPEAPLNYSRVAQLCRGQGGESSLSVSKWNTFLKAMEVC 259
Db 294 SVPGNGIDTHFELQDVFLM-NSKDP-KNPVIVGVFTTSSNIFKGSMDVRRV 351
Qy 260 S--DA-ATNKNFNRLQDVFLLDPGSGWRDTRVYVFSNPNW-Y--SAVCVLSLIDDKV 313
Db 352 FLGPYAHRDGPNYQGRVPYPRPGTCPSKTEGGFDSTKDLDDVITFGSRSHPAMYN 411
Qy 314 FRTS-SLK-G--YH--S--S-LPNRPRGKCLPDQ-QPI-PTETF--QV-A-DR-HPEVAQ 357
Db 412 PVFPINN-RPIMIKTDVNYQFTQIVVDRVDAEDGQ-YDVMFICTDVGIVLVKVVSPKETW 469
Qy 358 RVEPMGPKLTPFHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTIHKYVE-PGEQE 414
```

```
Db 470 HDLEEVLEEMTVFREPTTISAMELSTKQOQLYIGSTAGVAQLPLHRDIYKACAECL 529
Qy 415 HSFAFIME-IQFERRAAIQTMSLDAERRKLYSSQWESVQVPLDLCEVYGGGCHGCLM 473
Db 530 ARDPYCAWDGSSCSRYPFTAKRTRRQDIRNGDPLTHCSDLHDHNPSPLEERIIYGV 589
Qy 474 SRDPYCGWDQGRCSISYSS-RSVL-QSINPAEPHKECPN-PKPDK--AP-LQ-KVSLA- 525
Db 590 ENSSTFLECSPKSORALVY-WQ 610
Qy 526 PNSRYLSC-PMESRHATYSWR 546

RESULT 8
ENTRY I58169 #type fragment
TITLE semaphorin III - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
04-Sep-1998
ACCESSIONS I58169
REFERENCE #authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;
Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.
#journal Neuron (1995) 14:949-959
#title Semaphorin III can function as a selective chemorepellent to
pattern sensory projections in the spinal cord.
#cross-references MUID:95267432
#accession I58169
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type mRNA
#residues 1-666 #label RES
#cross-references GB:LA0484; NID:g703189; PID:g703190
GENETICS SemIII
#gene #superfamily semaphorin
CLASSIFICATION #length 666 #checksum 9654
SUMMARY
```

```
Query Match 16.0%; Score 662; DB 2; Length 666;
Best Local Similarity 30.7%; Pred.No. 9.47e-117;
Matches 141; Conservative 122; Mismatches 149; Indels 48; Gaps 37;

Db 60 RGKSPYDPKLLTASLLIDGELYSGTA-ANFMGRDAIFRTLGHHPITRTEQDHSRLNDP 118
Qy 135 RGYAPSPDENSLVFEDEVYSTIRKQYNGK-IPRERRIRGESELYTS--DIV-MQNP 190
Db 119 RFIASHLIPESDNEDDKVYFFERNATGGEHSGKATHARIQICKNDFGGHRSI-VNWK 177
Qy 191 QFIKATIVHQ-DQAYDDKIYFFERNDPKNPEAPLNYSRVAQLCRGQGGESSLSVSKW 249
Db 178 TTELKARLICSPGPNGLDTHFELQDVFLM-NSKDP-KNPVIVGVFTTSSNIFKGSVAVC 235
Qy 250 NTEFLKAMELCVS--DA-ATNKNFNRLQDVFLLDPGSGWRDTRVYVFSNPNW-Y--SAVC 303
Db 236 MYMSDVRVRLVLLGYAHRDGPNYQGRVPYPRPGTCPSKTEGGFDSTKDLDDVIT 295
Qy 304 VYSLGIDIKV---F--RTSS-LK--GYHSSLNPRGKCLPDQ-QPI-PTETF--QV-A 349
Db 296 FARSHPAMYNVPFINN-RPIMIKTDVNYQFTQIVVDRVDAEDGQ-YDVMFICTDVGIVL 353
Qy 350 -DR-HPEVAQRVEPMGPKLTPFHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTIH 405
Db 354 KVVSPKETWHDLEEVLEEMTVFREPTTISAMELSTKQOQLYIGSTAGVAQLPLHRCDI 413
Qy 406 KVEE-PGEQHSFAFIME-IQFERRAAIQTMSLDAERRKLYSSQWESVQVPLDLCEV 463
Db 414 YKACAECLARDPYCAWDGSSCSRYPFTAKRTRRQDIRNGDPLTHCSDLQHDHNP 473
Qy 464 YGGGCHGCLMSRDPYCGWDQGRCSISYSS-RSVL-QSINPAEPHKECPN-PKPDK--AP 518
Db 474 SLEBRIYGVENSSTFLECSPKSORALVY-WQFORNEDR 512
Qy 519 -LQ-KVSLA-PNSRYLSC-PMESRHATYSWR-HKENVEQ 553
```



```
ORGANISM      #formal_name vaccinia virus
DATE          20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
              09-Sep-1997

ACCESSIONS    S29921
REFERENCE      S29907
#authors      Amegadzie, B.Y.
#submission   submitted to the EMBL Data Library, January 1991
#accession    S29921
#status       Preliminary
#molecule_type DNA
#residues     1-441 #label AME
#cross-references EMBL:X57318; NID:g62239; PID:g62254
SUMMARY       #length 441 #molecular_weight 50185 #checksum 6034

Query Match   11.2%  Score 463; DB 2; Length 441;
Best Local Similarity 33.9%  Pred. No. 3.67e-72;
Matches       87; Conservative 57; Mismatches 96; Indels 17; Gaps 16;

Db 115 LVCGTNNGNPKWK-IDGSDPKHGRGYAPYQNSKVTIISHNGC-VLSDINISK-EG-I 170
QY 109 LACGTNARHPCSNLYNGTVVPLGEMRGYAPFSPDENSILVLFEGDEVYSTRKQYNGKI 168

Db 171 KWRFRDPGCGYDLYTADNVPKDG-LRGAFVDKGDYD-KVYILFTDTIGSKR--I-VK 225
QY 169 PRFRIRGE-S-ELYTSDTVNQNFQIKATIVHQDAYDDKIYFFREDNPKNPEAPLN 226

Db 226 IPYTAQCLNDEGSPSSLSHRWTFIKVELEC-DID-GRSY-R-Q-IIHSRTIKTD-ND 279
QY 227 VSRVAQLCRGOGGESSLSKWNFTLKAMLVCSDAATNKNFNRLQDVFLPDPGSGWRD 286

Db 280 TILVFFDPSYSKALCYNTIKQSFSTSKLEGYTKQLPSPASGICLPAGKVVPHTF 339
QY 287 TRYGVFSNPWNYSAVCVYSLGDIKDVFTSSLGKGYHSSLPNPRPGKCLPQQPIPTTF 346

Db 340 EVIEKYNVLDDIIRPLS 356
QY 347 QVADRHPVEAQRVPMG 363

RESULT 15
ENTRY      JH0798      #type complete
TITLE      fasciclin IV precursor - American bird grasshopper
ORGANISM   #formal_name Schistocerca americana #common_name American
           bird grasshopper
DATE       30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
           09-Sep-1997
ACCESSIONS JH0798
REFERENCE   JH0798
#authors   Kolodkin, A.; Matthes, D.J.; O'Connor, T.P.; Patel, N.H.;
           Admon, A.; Bentley, D.; Goodman, C.S.
#journal   Neuron (1992) 9:831-845
#title     Fasciclin IV: Sequence, expression, and function during
           growth cone guidance in the grasshopper embryo.
#accession JH0798
#molecule_type mRNA
#residues  1-730 #label KOL
#cross-references GB:L00709; NID:g160844; PID:g160845
#experimental_source embryo
COMMENT    This protein plays a role in growth cone guidance in the developing
           central nervous system.
KEYWORDS   glycoprotein; transmembrane protein
FEATURE
1-22       #domain signal sequence #status predicted #label SIG\
23-730     #product fasciclin IV #status predicted #label MAT\
23-627     #domain extracellular #status predicted #label EXT\
628-652    #domain transmembrane #status predicted #label TMM\
653-730    #domain intracellular #status predicted #label INT\
44,71,163,267,360,
539        #binding_site carbohydrate (Asn) (covalent) #status
           predicted
SUMMARY    #length 730 #molecular_weight 81214 #checksum 5881

Query Match   11.2%  Score 461; DB 2; Length 730;
```

```
Best Local Similarity 28.0%  Pred. No. 1.01e-71;
Matches       143; Conservative 130; Mismatches 182; Indels 56; Gaps 46;

Db 32 QFGEERVORFLGNESHKDKHFKLEKDHNSLLVGARNIVYNIISRLDTFTFTQRIEWHSSG 91
QY 29 HVGQDRVD-FQOTEPTHTVLFH--EPGSSSVWVGGRGVYLFDFPEGKN-ASVRTV-NI-G 82

Db 92 AHRELCLYKGSDEDDCONYIRVLAKIDDDRVLCGTNAYKPLCRHYALKDGDYVVEREYE 151
QY 83 STKGSC-L-DKR--DCENYITLLRSEGL-LLACGTNARHPSCWN--LVNGT-VVPLG-E 133

Db 152 GRGCPDPDPHNSHAIYSEGLYSATV-ADFSGTDLP-LI-YRG-P-LRTERSDL-K-QLN 204
QY 134 MRGYAPFSPDENSILVLFEGDEVYSTIRKQYNGKIPFRIRGESELYTSDTVNQNFQFI 193

Db 205 APNFVN-TMEYNDPIFFFTFRETAVEYINCGKAIY-SRVARVCKHDKGPHOFG-DRWTSF 261
QY 194 KATIVHQDAYDDKIYFFREDNPD-KNPEAPLNVSRAQLCRGDOGGESSLVSXWNTF 252

Db 262 LKSELNCSVPGDYPPFYFNEIQSTSDIIEGNYGGQ-VEKLIYGVFTTPVNSIGGSACAFS 320
QY 253 LKAMLVCSDAATNK-NFNRLQDVF-LLPDP-SGQMRDTRVYGVFSNPNW-Y--SAVCYIS 306

Db 321 MKSILESFDGPFKEQETMNSNMLAVPSLKVPEPRPGOCVNDSTRTPDVSVNFVKSHLMD 380
QY 307 L-G--D-IDKVFRT-SLKG-YHS--SL--PNRPGKCLPQQPIPTFTFOVADRHPEVA 356

Db 381 EAVPAFTT-RPILIRISLOVRYFTKIAVDQQVTPDGPDKAYDVLFTIGTDGKVIKALNSASF 439
QY 357 QRVEPMGPLKTPLFH-S-KYHYQKAV-HRMOASHGETFHVLYLTDRGTIHKVVEPG-- 411

Db 440 DSSDTVDSVVEELQVLPPGPVKNLYVVRMDGDDSKLVVVSDDDEILAIAKLRHCGSKIT 499
QY 412 EQEHSFAFNIME-IQPFRAAAAIQTM---SLDAERRKLYVSSOWEVSQVPLDLCEVYG-G 466

Db 500 NCRECVSLQDPYCAWMDNVELKCTAVGSPDWS 530
QY 467 GCHGCLMSRDPYCGWDQ-G-RCISIIYSERS 495
```

Search completed: Thu Jul 8 20:01:51 1999
Job time : 31 secs.

THIS PAGE BLANK (USPTO)

W P E R L H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 19:59:29 1999; MasPar time 17.63 Seconds
Tabular output not generated. 891.316 Million cell updates/sec

Title: >US-09-041-236-2
Description: (6-561) from US09041236.pep (44 of 45)
Perfect Score: 4130
Sequence: 1 WAAASAAQHLRSGPRIFAV.....TYSWRHKNVEQCEPGHQS 556

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28269293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 50.023; Variance 75.833; scale 0.660

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	470	11.4	403	1	VA39_VACCC	3.94e-86
2	463	11.2	441	1	PROTEIN A39.	2.47e-84
3	163	3.9	1871	1	SEX_HUMAN	8.57e-13
4	105	2.5	227	1	TRANSMEMBRANE PROTEIN	3.52e-02
5	102	2.5	275	1	HYPOTHETICAL PROTEIN H	1.03e-01
6	99	2.4	275	1	T-CELL ECTO-ADP-RIBOSY	2.94e-01
7	99	2.4	460	1	T-CELL ECTO-ADP-RIBOSY	2.94e-01
8	99	2.4	591	1	ELONGATION FACTOR 1-AL	2.94e-01
9	99	2.4	591	1	VP2_SALCH	2.94e-01
10	99	2.4	593	1	VP2_SALCH	2.94e-01
11	99	2.4	593	1	VP2_SALCH	2.94e-01
12	99	2.4	916	1	VP2_SALCH	2.94e-01
13	98	2.4	1132	1	VP2_SALCH	2.94e-01
14	95	2.3	255	1	HA21_HUMAN	1.14e+00
15	94	2.3	329	1	HA21_HUMAN	1.14e+00
16	93	2.3	380	1	HA21_HUMAN	1.14e+00
17	95	2.3	518	1	HA21_HUMAN	1.14e+00
18	95	2.3	518	1	HA21_HUMAN	1.14e+00
19	95	2.3	683	1	HA21_HUMAN	1.14e+00
20	93	2.3	827	1	HA21_HUMAN	1.14e+00
21	94	2.3	947	1	HA21_HUMAN	1.14e+00
22	94	2.3	959	1	HA21_HUMAN	1.14e+00
23	93	2.3	1029	1	HA21_HUMAN	1.14e+00

24	94	2.3	1069	1	ENTK_MOUSE	1.59e+00
25	94	2.3	1115	1	SERINE/THREONINE-PROTE	1.59e+00
26	94	2.3	1200	1	PUTATIVE PRE-MRNA SPLIC	1.59e+00
27	95	2.3	1390	1	DDX8_CABEL	1.14e+00
28	90	2.2	119	1	HEPATOCYTE GROWTH FACT	5.78e+00
29	90	2.2	213	1	50S RIBOSOMAL PROTEIN	5.78e+00
30	90	2.2	213	1	AMEX_BOVIN	5.78e+00
31	90	2.2	254	1	AMELOGENIN, CLASS I PR	5.78e+00
32	91	2.2	348	1	HLA CLASS II HISTOCOMP	4.21e+00
33	90	2.2	356	1	DDL_ENTFA	4.21e+00
34	90	2.2	380	1	D-ALANINE--D-ALANINE L	4.21e+00
35	89	2.2	437	1	VP39_NPVLD	7.91e+00
36	92	2.2	442	1	GBB_MAIZE	7.91e+00
37	92	2.2	460	1	NIFS PROTEIN HOMOLOG.	7.91e+00
38	91	2.2	498	1	O-ANTIGEN EXPORT SYSTE	3.05e+00
39	91	2.2	598	1	3-OXOACYL-ACYL-CARRIE	3.05e+00
40	89	2.2	598	1	ELONGATION FACTOR 1-AL	4.21e+00
41	91	2.2	707	1	PSEUDOLYSIN PRECURSOR	4.21e+00
42	92	2.2	1132	1	IMMEDIATE-EARLY RESPON	4.21e+00
43	92	2.2	1663	1	REGENERATING LIVER NUC	7.91e+00
44	91	2.2	1666	1	GCN5_HCVVA	4.21e+00
45	90	2.2	1874	1	MAJOR DNA-BINDING PROT	3.05e+00

ALIGNMENTS

RESULT	ID	VA39_VACCC	STANDARD;	PRT;	403 AA.
AC	P21052;				
DT	01-FEB-1991 (REL. 17, CREATED)				
DT	01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)				
DT	01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)				
DE	PROTEIN A39.				
GN	A39R.				
OS	VACCINIA VIRUS (STRAIN COPENHAGEN)				
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;				
OC	ORTHOPOXVIRUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 91021027.				
RA	GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,				
RA	PAOLETTI E.;				
RT	"The complete DNA sequence of vaccinia virus."				
RL	VIROLOGY 179:247-266(1990).				
RN	[2]				
RP	COMPLETE GENOME.				
RA	GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,				
RA	PAOLETTI E.;				
RL	VIROLOGY 179:517-563(1990).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; M35027; G335517; -				
DR	PIR; E42521; E42521.				
SQ	SEQUENCE 403 AA; 45741 MW; FE0DB9AB CRC32;				
	Query Match 11.4%; Score 470; DB 1; Length 403;				
	Best Local Similarity 33.5%; Pred. No. 3.94e-86;				
	Matches 86; Conservative 59; Mismatches 95; Indels 17; Gaps 16;				
Db	77 LVCGTNNGNPKWK-IDGSDDPKHRGYPAYONKSVTIISYN-ECVLSIDINTSK-BG-I 132				
Qy	109 LAGCTNHPSCNVLNVTGTVPLGEMRGYAFSPDENSELVLFEGDEVYSTIRKQYNGKI 168				
Db	133 KRWRRFDGPGCYDLYTADNVIPKDG-LRGAFAVDKDGTYD-KVYILFTDTIGSKR--I-VK 187				
Qy	169 PRFRIRGE-S-ELYTSDTYMNPQFIKATIVHODQAYDDKIYFFREDNPKNAPLN 226				

Qy 131 LGE-MRGYAPSPDENSILVLFEGDEVSTIRKQYNGK-IPFRIRIRGESE-LYTSOTVM 187
Db 252 ES 253
Qy 188 QN 189
::
RESULT 6
ID NRT1_RAT STANDARD; PRT; 275 AA.
AC P17982;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 PRECURSOR (EC 2.4.2.31) (T-CELL
NAD(P)(+)-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL MONO(ADP-
RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN
RT6.1).
GN ART2A OR RT6-A.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN:LEWIS A;
RX MEDLINE; 90192088.
RA HAAG F., KOCH F., THIELE H.-G.;
RT "Nucleotide and deduced amino acid sequence of the rat T-cell
alloantigen RT6.1";
RL NUCLEIC ACIDS RES. 18:1047-1047(1990).
RN [2]
RP MUTAGENESIS OF GLN-207.
RA MEDLINE; 96275529.
RA MAEHAMA T., HOSHINO S.-I., KATADA T.;
RT "Increase in ADP-ribosyltransferase activity of rat T lymphocyte
alloantigen RT6.1 by a single amino acid mutation.";
RL FEBS LETT. 388:189-191(1996).
CC -1- FUNCTION: HAS NAD+ GLYCOHYDROLASE ACTIVITY AND EXTREMELY LOW ADP-
CC -1- RIBOSYLTRANSFERASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +
CC N2-(ADP-D-RIBOSYL)-L-ARGININE.
CC -1- CATALYTIC ACTIVITY: NAD(+) + H2O = NICOTINAMIDE + ADP-RIBOSE.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.
CC -1- PFM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52082; G57168; -.
CC DR EMBL; M31138; G206804; -.
CC PIR; S08464; S08464.
CC PROSITE; PS01291; ART; 1.
CC PFAM; PF01129; ART; 1.
CC TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;
CC T-CELL DIFFERENTIATION; GPI-ANCHOR.
CC SIGNAL 1 20
FT CHAIN 21 246 T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.
FT PROPEP 247 275 HYDROPHOBIC, REMOVED DURING MATURATION
FT (BY SIMILARITY).
FT LIPID 246 246 GPI-ANCHOR (BY SIMILARITY).
FT ACT_SITE 209 209 BY SIMILARITY.
FT CARBOHYD 58 58 POTENTIAL.
FT MUTAGEN 207 207 Q->E: INCREASED ADP-RIBOSYLTRANSFERASE
FT ACTIVITY.
SQ SEQUENCE 275 AA; 31388 MW; 523B1A84 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 275;
Best Local Similarity 30.6%; Pred. No. 2.94e-01;
Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;
Db 194 LGYVIKEFS-FYPDQEE-VLIPGYEYQKVRTQGYNEIFLDSPKRSKNVNCILYSSAGTR 251
Qy 131 LGE-MRGYAPSPDENSILVLFEGDEVSTIRKQYNGK-IPFRIRIRGESE-LYTSOTVM 187
::
Db 252 ES 253
Qy 188 QN 189
::
RESULT 7
ID EFLA_TIRE STANDARD; PRT; 460 AA.
AC P34825;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 13-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEFL.
OS TRICHODERMA RESEI (HYPOCREA JECORINA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN:QM9414 / RUTC-30;
RX MEDLINE; 94124021.
RA NAKART T., ALATALO E., PENTTILA M.;
RT "Isolation of Trichoderma reesei genes highly expressed on glucose-
RT containing media: characterization of the tef1 gene encoding
RT translation elongation factor 1 alpha.";
RL GENE 136:313-318(1993).
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z33012; G312887; -.
CC PIR; S35772; S35772.
CC PROSITE; PS00301; EFATOR_GTP; 1.
CC PFAM; PF00009; GTP_EFTU; 1.
CC HSP; P07157; 1AIP.
CC ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.
KW NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 92 96 GTP (BY SIMILARITY).
FT NP_BIND 154 157 GTP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 49830 MW; B9ABAB2 CRC32;
Query Match 2.4%; Score 99; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 2.94e-01;
Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;
Db 210 YKGWEKETKAGKTG-KTLLFADSDIEPKR-PTDKPLRLPQDV 252
Qy 478 YCGWDQRCISYSSERSVLSQINPAEPHKECPNPKPKAPQKV 522
::
RESULT 8
ID VRP2_SALTY STANDARD; PRT; 591 AA.
AC P21454;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
65 KD VIRULENCE PROTEIN.
GN MKAA.
OS SALMONELLA TYPHIMURIUM.
OG PLASMID 96 KB VIRULENCE PEX102.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-TML R66;
RX MEDLINE; 90136009.
RA TAIRA S., RHEN M.;
RT "Identification and genetic analysis of mkaA -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth."
RL MICROB. PATHOG. 7:165-173(1989).
RN [2]
RN SEQUENCE OF 1-10.
RX MEDLINE; 91244158.
RA TAIRA S., BAUMANN M., RIIKONEN P., SUKUPOLVI S., RHEN M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium."
RL FEMS MICROBIOL. LETT. 61:319-323(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15042; G47783; -.
DR PIR; A54540; A54540.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 POLY-PRO.
SQ SEQUENCE 591 AA; 65440 MW; B69F25C3 CRC32;
Query Match 2.4%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 2.94e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLGKTAARLSDPOAASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RCENYITLLRRSEGLACGTNARHPSCWNLVNGTVPLGEMRGYAPFSPDENSILVLF 151
Db 220 GNEAGDRSAMRYSKLV-QYGNATPAADLY 248
QY 152 GDEVSTIRKQYNGKIPRRIRGESELY 181
RESULT 9
ID VRP2_SALCH STANDARD; PRT; 591 AA.
AC P17450;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN (PROTEIN M2, IN M2A REGION).
OS SALMONELLA CHOLERAE-SUIS (SALMONELLA ENTERICA).
OG PLASMID PKDSC50.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-RF-1;
RX MEDLINE; 90245675.
RA MATSUI H.;
RT "Nucleotide sequences of genes encoding 32 kDa and 70 kDa
RT polypeptides in mba region of the virulence plasmid, pKDC50, of

Salmonella choleraesuis.";
RL NUCLEIC ACIDS RES. 18:2181-2181(1990).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52035; G46898; -.
DR PIR; S09498; S09498.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 POLY-PRO.
SQ SEQUENCE 591 AA; 65341 MW; BEA4A39A CRC32;
Query Match 2.4%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 2.94e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLGKTAARLSDPOAASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RCENYITLLRRSEGLACGTNARHPSCWNLVNGTVPLGEMRGYAPFSPDENSILVLF 151
Db 220 GNEAGDRSAMRYSKLV-QYGNATPAADLY 248
QY 152 GDEVSTIRKQYNGKIPRRIRGESELY 181
RESULT 10
ID VRP2_SALEN STANDARD; PRT; 591 AA.
AC P55220;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN SPVB.
OS SALMONELLA ENTERITIDIS.
OG PLASMID PNL2001.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=AL1190;
RX MEDLINE; 94362897.
RA SUZUKI S., KOMASE K., MATSUI H., ABE A., KAWAHARA K., TAMURA Y.,
RA KIJIMA M., DANBARA H., NAKAMURA M., SATO S.;
RT "Virulence region of plasmid PNL2001 of Salmonella enteritidis.";
RL MICROBIOLOGY 140:1307-1318(1994).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14490; G517164; -.
KW PLASMID; VIRULENCE.
FT DOMAIN 367 POLY-PRO.
SQ SEQUENCE 591 AA; 65350 MW; 9C5B2EE3 CRC32;
Query Match 2.4%; Score 99; DB 1; Length 591;


```

RN  SEQUENCE FROM N.A.
RX  MEDLINE; 95266321.
RA  COMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA  MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;
RT  "The DNA sequence of human herpesvirus-6: structure, coding content,
RL  and genome evolution.";
RL  VIROLOGY 209:29-51(1995).
CC  -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC  REPLICATION.
CC  -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC  -!- SIMILARITY: TO OTHER HERPESVIRUSES' DNA-BINDING PROTEIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; X92436; G1044871; --
CC  EMBL; X83413; G854020; --
DR  PFAM; PF00747; Viral_DNA_bp; 1.
KW  DNA-BINDING; DNA REPLICATION; ZINC-FINGER; NUCLEAR PROTEIN.
FT  ZN.FING 459 475 C4-TYPE.
SQ  SEQUENCE 1132 AA; 127762 MW; DB699A02 CRC32;

Query Match 2.4%; Score 98; DB 1; Length 1132;
Best Local Similarity 28.2%; Pred. No. 4.15e-01;
Matches 24; Conservative 18; Mismatches 37; Indels 6; Gaps 6;

Db 435 WNLNMSVYNAGNATEI-YNHLVNCANLCEFCGKC-C-OSCGTAMRVGTRLPAP 491
QY 297 WNYSAVCYSLGDIDKVFRTSLKGHSLPNRPGKCLPDQOPIPTETFOVADRHPVA 356

Db 492 KNVK-KEPLVMSMF-SRY-YAEVDI 513
QY 357 QRVEPMPGLKTFPHSKHYQKAV 381

RESULT 14
ID HA21 HUMAN STANDARD; PRT; 255 AA.
AC P01908.
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84168117.
RA AUFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPES D.,
RA STROMINGER J.L.;
RT "Isotypic and allotypic variation of human class II
RT histocompatibility antigen alpha-chain genes.";
RL NATURE 308:327-333(1984).
RN [2]
RP SEQUENCE OF 40-255 FROM N.A.
RX MEDLINE; 83065171.
RA AUFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT "cDNA clone for the heavy chain of the human B cell alloantigen DC1:
RT strong sequence homology to the HLA-DR heavy chain.";
RL PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```

CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; J00199; --; NOT_ANNOTATED_CDS.
CC  PIR; A02212; HLHUDC.
DR  PROSITE; PS00290; IG_MHC; 1.
DR  PFAM; PF00047; Ig; 1.
DR  PFAM; PF00993; MHC_II_alpha; 1.
DR  HSP; P01910; IIAK.
KW  MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT  SIGNAL 1 23
FT  CHAIN 24 255 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT  DQ(1) ALPHA CHAIN
FT  DOMAIN 24 110 EXTRACELLULAR ALPHA-1.
FT  DOMAIN 111 204 EXTRACELLULAR ALPHA-2.
FT  DOMAIN 205 217 CONNECTING PEPTIDE.
FT  TRANSMEM 218 240
FT  DOMAIN 241 255 CYTOPLASMIC TAIL.
FT  DISULFID 133 189 BY SIMILARITY.
FT  CARBOHYD 104 104 POTENTIAL.
FT  CARBOHYD 144 144 POTENTIAL.
SQ  SEQUENCE 255 AA; 28105 MW; 5C46DD59 CRC32;

Query Match 2.3%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1.14e+00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;

Db 25 DIVADVAQLG-VNLYQSGPSGQYSHFDDGDEEYFVLERKETVQLPFRFR 78
QY 122 NLVNGTVVPLGMRGYAFSPDENSILVLFEGDEV-YSTIRKQYNGKIPFRFR 175

RESULT 15
ID RBSR_ECOLI STANDARD; PRT; 329 AA.
AC P25551.
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOSE OPERON REPRESSOR.
GN RBSR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6 AND 250-256.
RX STRAIN-K12.
RX MEDLINE; 93278299.
RA MAUZY C.A., HERMUDSON M.A.;
RT "Structural and functional analyses of the repressor, RbsR, of the
RT ribose operon of Escherichia coli.";
RL PROTEIN SCI. 1:831-842(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE; 93315143.
RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL GENOMICS 16:551-561(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RX MEDLINE; 93278300.
RA MAUZY C.A., HERMUDSON M.A.;
RT "Structural homology between rbs repressor and ribose binding protein
RT implies functional similarity.";
RL PROTEIN SCI. 1:843-849(1992).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RBSDACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING
```

Search completed: Thu Jul 8 19:59:59 1999
Job time : 30 secs.

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 20:00:16 1999; MasPar time 36.84 Seconds

Tabular output not generated. 823.798 Million cell updates/sec

Title: >US-09-041-236-2
Description: (6-561) from US09041236.pep (44 of 45)
Perfect Score: 4130

Sequence: 1 WAAASAQHLRSGPRIFAV.....TYSWRHKNVEQSCPEGHQS 556

Scoring table:
GAP 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_plant 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 48.573; Variance 74.876; scale 0.649

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	4130	100.0	666	4	SEMAPHORIN L.	0.00e+00
2	2372	57.4	293	11	SEMAPHORIN L. (FRAGMENT)	0.00e+00
3	1809	43.8	653	14	SIMILAR TO GENBANK ACC	0.00e+00
4	737	17.8	748	11	SEMAPHORIN A PRECURSOR	1.63e-154
5	725	17.6	751	11	SEMAPHORIN E PRECURSOR	2.38e-151
6	716	17.3	749	4	SEMAPHORIN V.	5.59e-149
7	716	17.3	750	4	SEMAPHORIN V.	5.59e-149
8	710	17.2	751	4	SEMAPHORIN E.	2.12e-147
9	705	17.1	751	13	COLLAPLIN 3.	4.38e-146
10	688	16.7	753	4	SEMAPHORIN III FAMILY	1.28e-141
11	682	16.5	754	11	SEMAPHORIN IV ISOFORM	4.82e-140
12	681	16.5	772	13	COLLAPLIN.	8.82e-140
13	682	16.5	785	11	SEMAPHORIN IV ISOFORM	8.82e-140
14	681	16.5	785	4	SEMAPHORIN IV.	8.82e-140
15	679	16.4	785	4	SEMAPHORIN.	2.95e-139
16	669	16.2	772	11	SEMAPHORIN D PRECURSOR	1.23e-136
17	664	16.1	772	11	SEMAPHORIN III/COLLAPS	2.52e-135
18	656	15.9	771	4	SEMAPHORIN-III.	3.12e-133
19	618	15.0	782	11	SEMAPHORIN C (SEM C)	2.58e-123
20	614	14.9	775	11	SEMAPHORIN H.	2.84e-122

21	599	14.5	861	11	009126	SEMAPHORIN J (SEMAPHOR	2.25e-118
22	592	14.3	775	4	015041	KIAA0331.	1.48e-116
23	568	13.8	294	13	090664	COLLAPLIN-3 (FRAGMENT)	2.42e-110
24	569	13.8	761	13	090663	COLLAPLIN-2.	1.33e-110
25	572	13.8	785	13	042237	COLLAPLIN 5.	2.24e-111
26	538	13.0	862	4	092854	SEMAPHORIN.	1.32e-102
27	522	12.6	834	11	064151	SEMAPHORIN I (M-SEMA F	1.69e-98
28	511	12.4	299	13	090666	COLLAPLIN-5 (FRAGMENT)	1.11e-95
29	473	11.5	295	13	090665	COLLAPLIN-4 (FRAGMENT)	5.39e-86
30	474	11.5	760	11	062178	SEMAPHORIN B PRECURSOR	3.00e-86
31	461	11.2	730	5	026473	FASCICLIN IV.	5.93e-83
32	448	10.8	1074	4	013591	SEMAPHORIN F HOMOLOG.	1.14e-79
33	441	10.7	1077	11	062217	SEMAPHORIN F PRECURSOR	6.64e-78
34	416	10.1	1093	11	060519	SEMAPHORIN G PRECURSOR	1.25e-71
35	392	9.5	888	11	035464	SEMAPHORIN VIA.	1.18e-65
36	390	9.4	494	4	060408	SEMAPHORIN F (FRAGMENT	3.71e-65
37	378	9.2	712	5	026972	SEMAPHORIN-I PRECURSOR	3.45e-62
38	367	8.9	562	5	017330	CESEMA.	1.77e-59
39	367	8.9	771	5	024322	SEMAPHORIN-I.	1.77e-59
40	362	8.8	706	5	024323	SEMAPHORIN-III.	2.98e-58
41	341	8.3	284	11	054948	SEMAPHORIN IV HOMOLOG	4.01e-53
42	316	7.7	886	11	034951	SEMAPHORIN N (SEMAPHOR	4.40e-47
43	316	7.7	887	11	070141	SEMAPHORIN Z.	4.40e-47
44	283	6.9	770	5	044253	SIMILAR TO SEMAPHORIN-	3.08e-39
45	204	4.9	1963	4	075051	KIAA0463 PROTEIN (FRAG	2.77e-21

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	666 AA.
ID	075326			
AC	075326;			
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	SEMAPHORIN L.			
GN	SEMAL			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC	CATARRHINI; HOMINIDAE; HOMO.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 9839619.			
RA	LANG E C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSSER A.;			
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA			
RT	viruses."			
RL	GENOMICS 51:340-350(1998).			
DR	EMBL; AF030698; G3523115; -.			
SQ	SEQUENCE 666 AA; 74823 MW; F546B39D CRC32;			

Query Match	100.0%;	Score 4130;	DB 4;	Length 666;
Best Local Similarity	100.0%;	Pred. No. 0.00e+00;		
Matches	556;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Db	38	WAAASAQHLRSGPRIFAVKMGHVQDRVDFGQTEPTVLFHEPGSSVWVGGRKYL	97	
Qy	6	WAAASAQHLRSGPRIFAVKMGHVQDRVDFGQTEPTVLFHEPGSSVWVGGRKYL	65	
Db	98	FDPEGKNASVTVNIGSTKSCLDKDCENVTILLRRSEGLACGTNARHPSCWNLVN	157	
Qy	66	FDPEGKNASVTVNIGSTKSCLDKDCENVTILLRRSEGLACGTNARHPSCWNLVN	125	
Db	158	GTVPPLGEMRGYAPSPDENSLVLFEGDEVYSTIRKQYNGKIPFRIRGESELYTSDT	217	
Qy	126	GTVPPLGEMRGYAPSPDENSLVLFEGDEVYSTIRKQYNGKIPFRIRGESELYTSDT	185	
Db	218	VQNQPFKATIHODQAYDKIYFFREDNPNKPEAPLNVSRVAQLCRDQGGESSLS	277	
Qy	186	VQNQPFKATIHODQAYDKIYFFREDNPNKPEAPLNVSRVAQLCRDQGGESSLS	245	
Db	278	VSKWNTFLKMLVCSDAATNKNFNLDQVLLPDPGQWRDTRVYGVFSPNPNYSACVY	337	
Qy	246	VSKWNTFLKMLVCSDAATNKNFNLDQVLLPDPGQWRDTRVYGVFSPNPNYSACVY	305	

Db	338	SLGIDKVFRTSSLKGYHSSLPNRPCKLPPDQOPIPTETQVADRHPVAQRVEPMGPL	397
Qy	306	SLGIDKVFRTSSLKGYHSSLPNRPCKLPPDQOPIPTETQVADRHPVAQRVEPMGPL	365
Db	398	KTPLFHSKYHOKVAVHRMOASHGTEHVLVLTDRGTIHKVVEPGOEHSFAFNIMEIQ	457
Qy	366	KTPLFHSKYHOKVAVHRMOASHGTEHVLVLTDRGTIHKVVEPGOEHSFAFNIMEIQ	425
Db	458	PFRAAAAIQTMSLDAERKLYVSSQWEVSQVPLDLCVYGGCHGCLMSRDPYCGWDOGR	517
Qy	426	PFRAAAAIQTMSLDAERKLYVSSQWEVSQVPLDLCVYGGCHGCLMSRDPYCGWDOGR	485
Db	518	CISYSSERSVLQINPAEPKPCPNKAPLOKQVSLAPNSRYLSCPMESRHATYSW	577
Qy	486	CISYSSERSVLQINPAEPKPCPNKAPLOKQVSLAPNSRYLSCPMESRHATYSW	545
Db	578	RHKNVQSCEPQHQS	593
Qy	546	RHKNVQSCEPQHQS	561
RESULT 2 PRELIMINARY; PRT; 393 AA.			
ID	O88371	AC	O88371;
DT	01-NOV-1998	(TREMBREL. 08, CREATED)	
DT	01-NOV-1998	(TREMBREL. 08, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998	(TREMBREL. 08, LAST ANNOTATION UPDATE)	
DE	SEMAPHORIN L (FRAGMENT).		
GN	SEWAL.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.		
RP	SEQUENCE FROM N.A.		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 98389619.		
RA	LANGHE C., LIEHR T., GOEN M., GERHART E., FLECKENSTEIN B., ENSER A.;		
RT	"New eukaryotic semaphorins with close homology to semaphorins of DNA		
RT	viruses".		
RL	GENOMICS 51:340-350(1998).		
DR	EMBL; AF030699; G3523117; -.		
FT	NON_TER	393	393
SQ	SEQUENCE	393 AA; 43909 MW; 7CB8193C CRC32;	
Query Match 57.4%; Score 2372; DB 11; Length 393;			
Best Local Similarity 90.8%; Pred. No. 0.00e+00;			
Matches 326; Conservative 14; Mismatches 15; Indels 4; Gaps 2;			
Db	38	WAAASAQGHRSRSPRISAVKNG---QDHVDFSQPEPHTVLFHEPGSFSVMVGGRGKYH	94
Qy	6	WAAASAQGHLSRGPRIFAVKGHVGDQVDFGQTEPHTVLFHEPGSSVMVGGRGKYV	65
Db	95	FNPEGKNASVRYTNIIGSTKGCODKCGNYITLLERRGNGLLVCGTNARKPSCNVLN	154
Qy	66	FDPEGKNASVRYTNIIGSTKGCODKCGNYITLLERRSEGLACGTNARHPSCNVLN	125
Db	155	DSVYMSLCEMKGYAPFPDENSLVFEDEYVSTIRKQYNGKIPRRFRIRGESELYTSD	214
Qy	126	GTVP-PLGEMRGYAPFPDENSLVFEDEYVSTIRKQYNGKIPRRFRIRGESELYTSD	184
Db	215	TVMQNPQFIKATIVHQDQAYDDKIYFFREDNPKNPEAPLNVSRAQLCRGQGGESSL	274
Qy	185	TVMQNPQFIKATIVHQDQAYDDKIYFFREDNPKNPEAPLNVSRAQLCRGQGGESSL	244
Db	275	SVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLPDPSGQWRDTRVYGFSPNPNYSACV	334
Qy	245	SVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLPDPSGQWRDTRVYGFSPNPNYSACV	304
Db	335	YSLGIDRVFTSSLKGYHMSLNPQCMCLPKKQPIPTETQVADSHPEVAQRVEPMG	393
Qy	305	YSLGIDRVFTSSLKGYHMSLNPQCMCLPKKQPIPTETQVADSHPEVAQRVEPMG	363

RESULT 3 PRELIMINARY; PRT; 653 AA.			
ID	O64906	AC	O64906;
DT	01-NOV-1996	(TREMBREL. 01, CREATED)	
DT	01-NOV-1996	(TREMBREL. 01, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998	(TREMBREL. 08, LAST ANNOTATION UPDATE)	
DE	SIMILAR TO GENBANK ACCESSION NUMBER L26081.		
OS	ALCELAAPHINE HERPESVIRUS 1.		
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;		
OC	GAMMAHERPESVIRINAE.		
RP	SEQUENCE FROM N.A.		
RP	SEQUENCE FROM N.A.		
RX	STRAIN=C500;		
RX	MEDLINE; 97201573.		
RA	ENSER A., FLECKENSTEIN B.;		
RT	"Alcelaphine herpesvirus type 1 has a semaphorin-like gene.;"		
RL	J. GEN. VIROL. 76:1063-1067(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RP	SEQUENCE FROM N.A.		
RX	STRAIN=C500;		
RX	MEDLINE; 97404659.		
RA	ENSER A., PFLANZ R., FLECKENSTEIN B.;		
RT	"Primary structure of the alcelaphine herpesvirus 1 genome.;"		
RL	J. VIROL. 71:6517-6525(1997).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RP	SEQUENCE FROM N.A.		
RX	STRAIN=C500;		
RA	ENSER A., PFLANZ R., FLECKENSTEIN B.;		
RL	SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.		
DR	EMBL; U18243; G1000717; -.		
DR	EMBL; AF003370; G2337970; -.		
SQ	SEQUENCE	653 AA; 73645 MW; CCCB53C9 CRC32;	
Query Match 43.8%; Score 1809; DB 14; Length 653;			
Best Local Similarity 47.18%; Pred. No. 0.00e+00;			
Matches 261; Conservative 98; Mismatches 183; Indels 12; Gaps 12;			
Db	62	SAITAAKSRFIDPRILVNLDTFGQHRF-FGPOEPHTVLFHLSNSDVYVGNNTIYLF	120
Qy	7	AAAASAQGHLSRGPRIFAVKGHVGDQVDFGQTEPHTVLFHEPGSSVMVGGRGKYV	66
Db	121	DFAHSSNASTALINITSNTHRLSSTCENFTILLNQDGLLACGTNSOKPSCW-LINN	179
Qy	67	DPEGKNASVRYTNIIGSTKGCODKCGNYITLLERRSEGLACGTNARHPSCNVLN	126
Db	180	LTTQFLGPKLGLAPFSPSSGNLVLFQDNDTYSTINLYKSLSGS-HKFERRIAGOVELYTSD	238
Qy	127	TVVP-LGEMRGYAPFPDENSLVFEDEYVSTIRK-QEYNGKIPRRFRIRGESELYTSD	184
Db	239	TAMHRPQFQVQATAVHKNESYDDKIYFFFOENSHSDFKQFPHTVPRYQGVCSDDGGESSL	298
Qy	185	TVMQNPQFIKATIVHQDQAYDDKIYFFREDNPKNPEAPLNVSRAQLCRGQGGESSL	244
Db	299	SVYKWTFLKARLACVDYDTGRIYNELQDIFITWOAPENSWEETLIYGLFLSPNFSACV	358
Qy	245	SVSKWNTFLKAMLVCSDAATNKNFNRLQDVFLPDPSGQWRDTRVYGFSPNPNYSACV	304
Db	359	FTVKDIDHVFKTSKLNKYNHKKLTPBPGCMKNHQPVTETFOVADRYPEVAPVYQKN	418
Qy	305	YSLGIDRVFTSSLKGYHSSLPNRPCKLPPDQOPIPTETQVADRHPEVAQRVEPMGP	364
Db	419	AMPEIIQSKYIYTKLAVRYVEYG-GVFWATIFYLTIKGTIHIYRYVEDSNSTALNILE	477
Qy	365	LKTPLFHSKYHOKVAVHRMQASHGETF-VLYLTDRGTIHKVVEPGOEHSFAFNIME	423
Db	478	INFQKAPATQNTILLONTNKLKLYNSEWESEVSEVPLDLCVYGNDCSCFNSRDPCLTWN	537
Qy	424	IQPFRAAAAIQTMSLDAERKLYVSSQWEVSQVPLDLCVYGGCHGCLMSRDPYCGWQ	483
Db	538	NTC-S-FK-QRVSVETGGPANRTLSEWCGDHYAPTVMKHQVSTPLLSNSYLSLCPAVSNHA	594
Qy	484	GRCSISYSSERSVLQINPAEPH-KE-CPNPKPKAPLQKVLAPNSRYLYLSCPMESRHA	541

[illegible]


```

01-MAY-1997 (TREMREL. 03, CREATED)
01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
SEMAPHORIN E. (HUMAN)
HOMO SAPIENS (HUMAN)
EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
CATARRHINI; HOMINIDAE; HOMO.
[1]
SEQUENCE FROM N.A.
RA YAMADA T., ENDO R., GOTOH M., HIROHASHI S.;
RT "Identification of semaphorin E as a non-MDR drug resistance gene of
human cancers."
RL PROC. NATL. ACAD. SCI. U.S.A. 94:14716-14718(1997).
DR EMBL; AB000220; D1033360; -.
SQ SEQUENCE 751 AA; 85207 MW; F89FB934 CRC32;

Query Match 17.2%; Score 710; DB 4; Length 751;
Best Local Similarity 33.6%; Pred. No. 2,12e-147;
Matches 151; Conservative 104; Mismatches 151; Indels 43; Gaps 30;

Db 167 SENPNVNTVSMINEELFSGMYI-DFMGTDAAIFSLTKRNAVRTDOHNSKWLSEPFYVD 225
QY 139 PESPENSLVLFEGDEVISTIRKQYNGK-IPFRIRGESELYTS--DTV-MONPOFIK 194
Db 226 AHVIPDGTDPNDAKYVFFFEKELTDNRSSTKQIHSMIARICPNDTGGLRSL-VNKKWTFEL 284
QY 195 ATIVHOD-QAYDDKIYFFREDNPDNPEAPLNVSRAQLCRGDQGESLSVSKNNTFL 253
Db 285 KARLVCSVDEDEGPTHFDELEDFVLL-ETDNP-RTTLVYGIFTSSSVFKGSACVYHL 342
QY 254 KAMLVCS--DA-ATKNFNRLQDVFLLPDPGQWRDTRVYGVFSPWN-Y--SAVCVYSL 307
Db 343 SDIQTVFNGPFAHKEGPNHQLISYQGRIPYPRPGTCGGAGFPNMRTTKEFFDDVVFIR 402
QY 308 GDIKVFRTS-SLK-G-----YHSSLNPRPGKLPDQ-QP-IPETET-F--QVAD--R 351
Db 403 NHPLMYNSIPIYHKRPLIVRI-GTDYKTKIAVDRNAADG-RYHVLFGTDGRTGVOKVY 460
QY 352 -HPEVAQRVEPM-G-PLKTPHSHKHYOKVAVHRMQASHGETFHVLYLTDRGTIHKVY 408
Db 461 VLPTNNSVSGELILEEVLKFNHAPITTKKISSKKOOLYSSNEGYSQVSLHRCYIYGA 520
QY 409 E-PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERRKLYVSSQWESVQVPLDLCEVYGG 467
Db 521 CADCCCLARDPYCAWDGSCSRFPYPTGKRSSRDQVRHGNPLTCRGFNLKAYRNAAEIYO 580
QY 468 CHGCLMSRDPYCGWDGRCISYISSE--RSVLQSIINPAEPHKECP--NPKPKAPLQKVS 523
Db 581 YGKNNNTTLECAPKSPQASIKWLLQDK 609
QY 524 LAP-NSRYILSCPMESRHATYSWR-HKEN 550

RESULT 9
ID O42236 PRELIMINARY; PRT; 751 AA.
AC O42236:
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DE COLLAPIN 3.
OS GALLUS GALLUS (CHICKEN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA FEINER L., KOPPEL A.M., KOBAYASHI H., RAPER J.A.;
RL NEURON 19:0-0(1997)
DR EMBL; AF022946; G2522204; -.
SQ SEQUENCE 751 AA; 85432 MW; 0B9BA2F7 CRC32;

Query Match 17.1%; Score 705; DB 13; Length 751;
Best Local Similarity 33.5%; Pred. No. 4,38e-146;
Matches 150; Conservative 107; Mismatches 150; Indels 41; Gaps 28;

01-MAY-1997 (TREMREL. 03, CREATED)
01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
SEMAPHORIN E. (HUMAN)
HOMO SAPIENS (HUMAN)
EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
CATARRHINI; HOMINIDAE; HOMO.
[1]
SEQUENCE FROM N.A.
RA YAMADA T., ENDO R., GOTOH M., HIROHASHI S.;
RT "Identification of semaphorin E as a non-MDR drug resistance gene of
human cancers."
RL PROC. NATL. ACAD. SCI. U.S.A. 94:14716-14718(1997).
DR EMBL; AB000220; D1033360; -.
SQ SEQUENCE 751 AA; 85207 MW; F89FB934 CRC32;

Query Match 17.2%; Score 710; DB 4; Length 751;
Best Local Similarity 33.6%; Pred. No. 2,12e-147;
Matches 151; Conservative 104; Mismatches 151; Indels 43; Gaps 30;

Db 167 SENPNVNTVSMINEELFSGMYI-DFMGTDAAIFSLTKRNAVRTDOHNSKWLSEPFYVD 225
QY 139 PESPENSLVLFEGDEVISTIRKQYNGK-IPFRIRGESELYTS--DTV-MONPOFIK 194
Db 226 AHVIPDGTDPNDAKYVFFFEKELTDNRSSTKQIHSMIARICPNDTGGLRSL-VNKKWTFEL 284
QY 195 ATIVHOD-QAYDDKIYFFREDNPDNPEAPLNVSRAQLCRGDQGESLSVSKNNTFL 253
Db 285 KARLVCSVDEDEGPTHFDELEDFVLL-ETDNP-RTTLVYGIFTSSSVFKGSACVYHL 342
QY 254 KAMLVCS--DA-ATKNFNRLQDVFLLPDPGQWRDTRVYGVFSPWN-Y--SAVCVYSL 307
Db 343 SDIQTVFNGPFAHKEGPNHQLISYQGRIPYPRPGTCGGAGFPNMRTTKEFFDDVVFIR 402
QY 308 GDIKVFRTS-SLK-G-----YHSSLNPRPGKLPDQ-QP-IPETET-F--QVAD--R 351
Db 403 NHPLMYNSIPIYHKRPLIVRI-GTDYKTKIAVDRNAADG-RYHVLFGTDGRTGVOKVY 460
QY 352 -HPEVAQRVEPM-G-PLKTPHSHKHYOKVAVHRMQASHGETFHVLYLTDRGTIHKVY 408
Db 461 VLPTNNSVSGELILEEVLKFNHAPITTKKISSKKOOLYSSNEGYSQVSLHRCYIYGA 520
QY 409 E-PGEQHSFAFNIMEIQPFRAAAIQTMSLDAERRKLYVSSQWESVQVPLDLCEVYGG 467
Db 521 CADCCCLARDPYCAWDGSCSRFPYPTGKRSSRDQVRHGNPLTCRGFNLKAYRNAAEIYO 580
QY 468 CHGCLMSRDPYCGWDGRCISYISSE--RSVLQSIINPAEPHKECP--NPKPKAPLQKVS 523
Db 581 YGKNNNTTLECAPKSPQASIKWLLQDK 609
QY 524 LAP-NSRYILSCPMESRHATYSWR-HKEN 550

RESULT 10
ID Q13372 PRELIMINARY; PRT; 753 AA.
AC Q13372:
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN III FAMILY HOMOLOG.
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96230324.
RA XIANG R.H., HENSEL C.H., GARCIA D.K., CARLSON H.C., KOK K., DALY M.C.,
RA KACHAR K., DEN BERG A., VELDHOIS P., BUYS C.H., NAYLOR S.L.;
RT "Isolation of the human semaphorin iii/F gene (SEMA3F) at chromosome
3p21, a region deleted in lung cancer."
RL GENOMICS 32:39-48(1996).
DR EMBL; U38276; G1061351; -.
DR PFAM; PF00047; 19; 1.
SQ SEQUENCE 753 AA; 84941 MW; BECBEBB0 CRC32;

Query Match 16.7%; Score 686; DB 4; Length 753;
Best Local Similarity 31.4%; Pred. No. 1,28e-141;
Matches 139; Conservative 121; Mismatches 140; Indels 42; Gaps 30;

Db 171 PYDPKLDTSALINEELYAGVYI-DFMGTDAAIFRTLKQTAMRTDQYNSRLNPDSPFIH 229
QY 139 PESPENSLVLFEGDEVISTIRKQYNGK-IPFRIRGESELYTS--DTV-MONPOFIK 194
Db 230 ABLIPDPAENDDKLYFFFRSAE-APQSPAYVARIGRICLNDGHCCL-VNKNSTFLK 287
QY 195 ATIVHODQAYDDKIYFFREDNPDNPEAPLNVSRAQLCRGDQGESLSVSKNNTFLK 254
Db 288 ARLVCSVPGDEGIETHFDELDQVVFV-QQTQDV-RNPVIYAVFTSSGVSRFGSAVYNSNA 345
QY 255 AMLVCS---DAATNKNFNRLQDVFLLPDPGQWRDTRVYGVFSPNP-WNY--SAVCVYSLG 308
```



```

RN      SEQUENCE FROM N.A.
RP      ECKHARDT F., MEYERHANS A.;
RT      "Molecular cloning and expression pattern of a murine semaphorin
RL      homologous to H-sema IV";
RL      SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR      EMBL; AF080090; G3377766; -.
SQ      SEQUENCE 785 AA; 88493 MW; 6FFCD8F5 CRC32;

Query Match      16.5%; Score 682; DB 11; Length 785;
Best Local Similarity 31.8%; Pred. No. 4.82e-140;
Matches 141; Conservative 119; Mismatches 140; Indels 43; Gaps 31;

Db      202 PYDPKLDTASALINEELYAGVYI-DFMGTDAAIFRLGKOTAMRTQOYNSRWLNDPSFIH 260
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      139 PFPSPDENSILFFGDEYVIRKOEYNGK-IPFRRIERGESELYTS-DTV-MONPOFIK 194
Db      261 AELIPDSAEKNDKLYFFFRERSAE-APQNPAYVAIGRICLNDNDGGHCLL-VNKHSTEL 318
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      195 ATIVHOD-QAYDDKIYFFREDNPDKNPEAPLNVSRVAQLCRDQDGESLSYSKWNTFL 253
Db      319 KARLVCSVPCEGDIETHFDLQDFV-QOTQDI-RNPVIYAVFTSSGVSFRGSAVCVISM 376
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      254 KAMLVCS--DAATNKNFRLQDVLLPDPDSGQWRDTRVYGVESNP-WNY--SAVCVYSL 307
Db      377 ADIRWVNGPFAHKEGPNYQWMPFSCKMPPRPRTGTCPGGTTFTSMKSTKDYDPDEVINFM 436
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      308 GDIDKVFRTS-SLK-G-YH-----SS-LPNPRGKCLPDQ-QP-I-PTETF--QVAD--R 351
Db      437 THPLMQAVYPLQRRPLVVRT-GAPRLTTVAVDQVDAADG-RYEVLFGLTDRGTGVOKVI 494
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      352 -HPEVAQRVEPMG--PLKTPLFHSHYQKVAVHRMQASHGETFHYLYLTTRDTGTHKVV 408
Db      495 VLPKDDQVEELMEEVEKEPAPVKYMTITSSKROOLYVASAVGTHLSLHRCQAYGAA 554
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      409 E-PGEQGHSAFNIQLPFRRAAQTMSLDAERKLYVSSQWVSPQLDCEVYGGG 467
Db      555 CADCCCLARDPYCAWDGQACSRYSATSSKRRSRRODVRRGNP IROCGRFNSNANKNAVESYQ 614
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      468 CHGLMSRDPYCGWDGRCIS-IYSSER-SVLQSIINPAEPHKPCP--NPKPDKAPLOKVS 523
Db      615 YGVAGSAAFLECPSPQATVKW 637
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      524 LA-PNSRYLSCPMESRHATYSW 545

RESULT 14
AC      Q13275 PRELIMINARY; PRT; 785 AA.
ID      Q13275; Q13274;
DT      01-NOV-1996 (TREMBLREL. 01, CREATED)
DT      01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT      01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE      SEMAPHORIN IV.
OS      HOMO SAPIENS (HUMAN).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC      CATARRHINI; HOMINIDAE; HOMO.
RN      [1]
RP      NELSON J., BIEWALD T.;
RL      SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RN      SEQUENCE OF 394-436 FROM N.A.
RC      TISSUE=PLACENTA;
RX      MEDLINE; 96210603.
RA      SKIDOO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
RA      ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT      "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT      cancer deletion region and demonstrate distinct expression
RT      patterns";
RL      PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR      EMBL; AC000063; G1669373; -.
DR      EMBL; U32172; G995788; -.
DR      EMBL; U32171; G995786; -.

```

WIRE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 20:05:59 1999; MasPar time 28.25 Seconds
414.834 Million cell updates/sec
Tabular output not generated.

Title: >US-09-041-236-2
Description: (55-605) from US09041236.pap (45 of 45)
Perfect Score: 4122
Sequence: 1 VVVGGRGKVVLFDFPEGKNA.....GSYFRAQHWWLLPEDGIMA 551
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.248; Variance 152.976; scale 0.237

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	661	16.0	771 13	R71380	Human semaphorin III	4.59e-51
2	620	15.0	477 13	R74175	Human collapsin.	4.69e-47
3	608	14.8	775 32	W63748	Human semaphorin.	6.95e-46
4	606	14.7	861 22	W17658	Mouse CD100 antigen.	1.09e-45
5	599	14.5	861 32	W58540	Human semaphorin.	5.24e-45
6	544	13.2	862 22	W17657	Human CD100 antigen.	1.16e-39
7	490	11.9	776 32	W51313	Rat semaphorin W.	1.90e-34
8	463	11.2	441 13	R71381	Vaccinia virus semaph	7.46e-32
9	445	10.8	730 13	R71379	Grasshopper semaphori	3.94e-30
10	428	10.4	587 32	W51314	Human semaphorin W.	1.65e-28
11	390	9.5	974 33	W64221	Human secreted protei	6.69e-25
12	367	8.9	650 13	R71382	Drosophila semaphorin	9.87e-23
13	368	8.9	712 13	R71384	Tribolium semaphorin	7.95e-23
14	362	8.8	724 13	R71383	Drosophila semaphorin	2.91e-22
15	352	8.5	930 32	W57260	Human semaphorin Y.	2.52e-21
16	325	7.9	888 25	W15857	Human semaphorin Z.	8.34e-19

Rat semaphorin Z. 5.70e-18
Rat semaphorin Y. 4.00e-16
Variola major virus s 7.29e-04
Partial BRCA2 cancer 1.70e-01
Protein transcribed f 2.37e+01
Taq DNA polymerase I 2.79e+01
Taq DNA polymerase I 3.87e+01
BRCA2 cancer suscepti 1.18e+02
Recombinant elastase. 8.63e+01
Partial BRCA2 cancer. 1.18e+02
Sequence of nuclear r 8.63e+01
E. coli HSP (dnaK). 1.01e+02
Mutant thermostable D 1.18e+02
Mutant thermostable D 1.18e+02
Mutant thermostable D 1.18e+02
Taq DNA polymerase I 7.36e+01
DNA-polymerase REM-T3 1.18e+02
T. aquaticus DNA-poly 1.18e+02
Purified native therm 1.18e+02
Taq DNA polymerase I 1.18e+02
Taq DNA-polymerase RE 1.18e+02
Cleavase DN nuclease. 1.18e+02
Cleavase DA nuclease. 1.18e+02
Cleavase DV nuclease. 1.18e+02
Clas II S-receptor ki 7.36e+01
S receptor kinase pro 7.36e+01
Rattus norvegicus cdo 1.01e+02
Human breast cancer s 1.18e+02
Human breast and ovar 1.18e+02

ALIGNMENTS

RESULT 1

ID R71380 standard; Protein; 771 AA.

AC R71380;

DT 21-NOV-1995 (first entry)

DE Human semaphorin III protein.

KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW variola major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection.
OS Homo sapiens.
PN W09507706-A.
PS 23-MAR-1995.

PD 23-MAR-1995.

PF 13-SEP-1994; U10151.

PR 13-SEP-1993; US-121713.

PA (REGC) UNIV CALIFORNIA.

PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;

PI O'Connor T;

DR WPI; 95-131177/17.

DR N-PSDB; Q87442.

PT New class of semaphorin peptide(s) and polypeptide(s) - are

potent modulators of nerve cell growth and regeneration

PS Example 2; Page 60-63; 101pp; English

CC The sequence of the human semaphorin III protein. The proteins

CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II
CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)
CC virus semaphorin IV (Q87447) genes were used to generate a series of
CC peptides (R70370-R70418), which retain semaphorin receptor binding
CC activity. The semaphorin derived or semaphorin receptor derived peptides
CC are potent modulators of nerve cell growth, immune responsiveness and
CC viral pathogenesis. They can be used in diagnosis and treatment of
CC neurological disease and neuro-regeneration, immune modulation and
CC diagnosis and treatment of viral and oncological infection and diseases.
SQ Sequence 771 AA;

Query Match 16.0%; Score 661; DB 13; Length 771;

Best Local Similarity 30.3%; Pred. No. 4.59e-51;

Matches 151; Conservative 136; Mismatches 174; Indels 47; Gaps 37;

Db 166 rgksydpklltasllldgelysgta-adfmgdrdfairtghhhpirtedqhsrwlndp 224

Db 343 vyhmssireafngpyahkegpeyhwslyegkvpyprpgscaskvnggkygttkdypddai 402
QY 304 VYSLGIDIKVFRFS-SLK-G--YHSSL-----PNPRGKCLPD-QQPI--PTTF-QVAD 350
Db 403 rfarmhplmqpikpyhkhkpylvtgdgkynrlqlavdrveaedgq-ydvlfigtdtgi 461
QY 351 RHPEVAQRV-EPMGPL-KTP-LFHS--KYHYQKVAVHRMQASHGETFHVLYLTDRGTIH 405
Db 462 kvitinyqetwemeevilleelqifkdpapiismieaskrqqlvigasavaqvrhfcdm 521
QY 406 KVVEPEGEQSHSFAFN-IME-IQPFRAAAIQTMSLDAERKLYVSSQWEVSQVPLDLCEV 463
Db 522 ygsacacclardpycawdgiscsryypt 550
QY 464 YGGGCHGCLMSRDPYCGWDQGRICISYSS 492

RESULT 4
ID W17658 standard; Protein; 861 AA.
AC W17658;
DT 24-JUL-1997 (first entry)
DE Mouse CD100 antigen.
KW CD100 antigen; semaphorin; leukocyte; B cell; T cell; lymphocyte;
KW vaccine.
OS Mus sp.
FH Key
FT peptide
FT /label= Sig_peptide
FT protein
FT /label= Mat_protein
FT domain
FT /label= Semaphorin_domain
FT domain
FT /label= Ig-like_domain
FT domain
FT /label= Stalk_domain
FT domain
FT /label= Transmembrane_domain
FT domain
FT /label= Cytoplasmic_domain
FT modified_site
FT /label= Phosphorylation
FT /note= "putative tyrosine phosphorylation site"
PN W09717368-A1.
PD 15-MAY-1997.
PF 12-NOV-1996; U18645.
PR 09-NOV-1995; US-556422.
PA (DAND) DANA FARBER CANCER INST.
PI Bousiotis V, Freeman GJ, Hall KT, Nadler LM, Schultze JL;
DR WPI: 97-280982/25.
DR N-PSDB; T60666.
PT Nucleic acid molecule encoding CD100 antigen - which stimulates
PT leukocyte response, e.g. B cell aggregation, differentiation,
PT survival and T cell proliferation
PS Example 8: Page 86-89; 135pp; English.
CC Mouse CD100 antigen (W17657) is a novel leukocyte semaphorin-like
CC protein that stimulates a leukocyte response, including B cell
CC aggregation, B cell differentiation, B cell survival and/or T cell
CC proliferation. Its amino acid sequence was deduced from a cDNA
CC clone (T60666) isolated from murine T cells. Human CD100 antigen
CC (W17657) has also been identified. CD100 polypeptides and fusion
CC proteins, nucleic acids, and host cells expressing CD100 can be
CC utilised in diagnostic and therapeutic methods involving modulation
CC of B and T cell responses, neuron axonal growth and immune cell-
CC nerve cell interaction.
SQ Sequence 861 AA;

Query Match 14.7%; Score 606; DB 22; Length 861;
Best Local Similarity 32.4%; Pred. No. 1.09e-45;
Matches 143; Conservative 106; Mismatches 145; Indels 47; Gaps 31;

Db 107 eclnyirvlqplsslslyvcgtnafptcdhnltsfkflgksedgkrcpdpahsyts 166

QY 93 DCENYITLLER-RSEGLLAGGINARHPS- -WNLVNGTVVPLGEM-RGYAPSPDENS 148
Db 167 vnmvgelysgts-ynflgsepilssnshspirteyaipwlnepsfvfadvikspdpge 225
QY 149 LFEQDEVYSTIRKQYNGKIPRFRIRGESELYTSDTV--MQNPQIKATIVHQ--D--Q 202
Db 226 geddkvyfftevsveyefvklmiprvarvckgdggirtiq-kkwtstflkarlicskp 284
QY 203 AYDDKIYFFREDNPKNPEAPLNVSRAQLCRGDGGESSUSVSKWNTFLKAMLVCSDA 262
Db 285 dsglvfnlqdvfvrwrp-gl-kepvfyavftqplnnvglsavcaytlatveavfargky 342
QY 263 ATKNFENRLQDVFLLPDPSSQWRDTRVYGVFSPNWN---SACVVSGLGIDIKVF-R--- 315
Db 343 mgsatveqshkvwryngvptprgacidsaaraanytsslnlpdktlqfvkdhplmd 402
QY 316 --TSSL-----K-G-YHSSLNPNRPGKCL-PDQOPIP-TETQVADRHP-E-V-AQR-VE- 360
Db 403 svtpidnrpklikkdvnyqtqdvdrtdqldgtfydvfmfistdrgalhhkaviltkev 460
QY 361 PMGPLKT-P-LFHSKYHYOKVAVHRMQASHGETFHVLYLTDRGTIHKVVEPEGEQSHSFA 418
Db 461 --ieetqlfridsepvtllllsskkgkfyagsgnsvvqaplafcekhgs-cedcvlar 517
QY 419 FNIMEIQPFRAAAIQTMSLDAER-RKL-YVSSQWEVSQVPLDLCEVYGGCHGCLMSRD 476
Db 518 pycawspaikacvtlhqeas 538
QY 477 PYCGWDQGR-CISYSSERS 495

RESULT 5
ID W58540 standard; Protein; 861 AA.
AC W58540;
DT 02-SEP-1998 (first entry)
DE Human semaphorin.
KW Human; semaphorin; diagnosis; nervous disease; immune disease;
KW nerve extension inhibiting activity.
OS Homo sapiens.
PN J10155490-A.
PD 16-JUN-1998.
PF 27-NOV-1996; 332900.
PR 27-NOV-1996; JP-332900.
PA (SUMO) SUMITOMO SEIYAKU KK.
DR WPI: 98-391044/34.
DR N-PSDB; V31121.
PT New human semaforin gene - useful in the diagnosis of nervous system
PT and immune disorders
PS Claim 1; Page 10-12; 15pp; Japanese.
CC The present sequence represents human semaphorin (translated from the
CC Japanese specification as semaforin). Semaphorin has nerve extension
CC inhibiting activity. The semaphorin gene is useful for the diagnosis,
CC treatment and researches on nervous diseases and immune diseases.
SQ Sequence 861 AA;

Query Match 14.5%; Score 599; DB 32; Length 861;
Best Local Similarity 32.4%; Pred. No. 5.24e-45;
Matches 143; Conservative 105; Mismatches 146; Indels 47; Gaps 31;

Db 107 eclnyirvlqplsslslyvcgtnafptcdhnltsfkflgksedgkrcpdpahsyts 166
QY 93 DCENYITLLER-RSEGLLAGGINARHPS- -WNLVNGTVVPLGEM-RGYAPSPDENS 148
Db 167 vnmvgelysgts-ynflgsepilssnshspirteyaipwlnepsfvfadvikspdpge 225
QY 149 LFEQDEVYSTIRKQYNGKIPRFRIRGESELYTSDTV--MQNPQIKATIVHQ--D--Q 202
Db 226 geddkvyfftevsveyefvklmiprvarvckgdggirtiq-kkwtstflkarlicskp 284
QY 203 AYDDKIYFFREDNPKNPEAPLNVSRAQLCRGDGGESSLSVSKWNTFLKAMLVCSDA 262
Db 285 dsglvfnlqdvfvrwrp-gl-kepvfyavftqplnnvglsavcaytlatveavfargky 342

QY 208 IYFFREDNPD-KNPEAPLNVSRAOLCRGQGGESSLSVSKWNTFLKAMLVCSDAATNK 266
Db 276 fyfneiqstsdilegnyggq-vekllygvfttpvnsiggsavcafsmksilesfdgpfke 334
QY 267 -NENRLQDVF-LLPDP-SGOWRDRTRYGVFSNPNW-Y--SACVCYSL-G-D-IDKVFRT 316
Db 335 qetmnsnwlavpslkyppeprpgcgvndrtldpvsfnvskshlmdeavpafft-rp11 393
QY 317 -SSLKG-YHS--SL--PNPRPGKCLPDQOPIPTETQVADRHPDEVAQRPVPMGLKPLF 370
Db 394 rlslyqfktiavdqgvrtpdgkaygvlfigtdgdgkvikalsasfddsdtdvdsvieel 453
QY 371 H-S-KYHOKVAV-HRMQASHGETFHLVLTDTDRGTHHKVVEPG--EQEHSFAFNIME-I 424
Db 454 qvlppgvvknlyvymdgdskslvvsddelaiklhr-cgskdincrcvslgqpyca 513
QY 425 QPFRRAAAIOTM--SLDAERRKLYVSSQWVEVSQVPLDCEVYG-GGCHGCLMSRDPYCG 480
Db 514 wdnelkctavspdw 530
QY 481 WDO-G-RCISIVSSERS 495

RESULT 10
ID W51314 standard; Protein: 587 AA.
AC W51314;
DE 08-SEP-1998 (first entry)
DT Human semaphorin W.
KW Human; semaphorin W; nerve extension inhibitor; antiallergic; anticancer;
KW Immunosuppressant; gene therapy; diagnosis; research reagent.
OS Homo sapiens.
PN W09815628-A1.
PD 16-APR-1998.
PF 03-OCT-1997; J03549.
PR 09-OCT-1996; JP-287636.
PA (SUMU) SUMITOMO PHARM CO LTD.
PI Kikuchi K, Kimura T;
DR WPI: 98-261015/23.
DR N-P5DB: V07281, V07282.
DR Nerve extension inhibitor protein semaphorin W - is useful as
PT therapeutic drug and diagnostic and research reagent
PS Example 4; Page 69-72; 50pp; Japanese.
CC The present sequence represents human semaphorin W. Semaphorin W and
CC its derivatives are nerve extension inhibitors which are useful as
CC antiallergic, immunosuppressant and anticancer agents. The DNA
CC encoding semaphorin W can also be used in gene therapy, e.g. using
CC a viral vector. The proteins, peptides, DNA and antibodies which
CC recognise the protein or peptides, can be used as diagnostic or
CC research reagents. Semaphorin W can be used as a screen for
CC semaphorin W antagonists with possible therapeutic use.
SQ Sequence 587 AA;

Query Match 10.4%; Score 428; DB 32; Length 587;
Best Local Similarity 28.9%; Pred. No. 1.65e-28;
Matches 108; Conservative 89; Mismatches 145; Indels 32; Gaps 22;

Db 57 dedgddelyffctetsratdsyerikvprvarvcagdlggrktlg-grwtftlkadllcp 115
QY 201 DQAYDDKIYFFREDPNKNPEAPLNVSRAOLCRGQGGESSLSVSKWNTFLKAMLVCS 260
Db 116 gpehgrassvldgavlrpelga-gtptifygiffssqwegatisavcafrpqdirtvlnp 174
QY 261 DAATKNFNRLQDVFLLPDPGQWRDRTRYGVFSNPNW---YSACVCYSLGDDIKVFRS 317
Db 175 frelkhdcnrglpvvdndvpqprgpcitnnmklrhfgsslsldrvlflfirdhplmdrp 234
QY 318 --SLK--GYHSSUP---N---PRPGKCLPDQOPIP--TETQVADRHPD-V-AQR-VE-P 361
Db 235 vfpadghplvttdtaylrvvahrvtlsqgkeydvlylgtedghllhravrigaqls-vl 292
QY 362 MGPLKT-PLFHSK-YHYQKVAHVMQASHGETFHLVLTDRGTHHKVVEPGQEHSFAF 419
Db 293 edlal--fpepqpvenmkl-yhsw-llvgsrtvtqnttcnrglqs-csecilagdpvc 347

QY 420 NTMEIQPFRRAAAIOTMSLDAERRKLYVSSQWVEVSQVPLDCEVYGGCHGCLMSRDPYC 479
Db 348 awsfldcevhagehrglvqdieadvsslcpcpkegperpvpvfevptaaahvlpccps 407
QY 480 GWD-Q-GRCISIVSSERSVLQSLNPAEPHKECPNPKAPLOKVLAPNSRYLSCPME 537
Db 408 sawascvwhqpsgv 421
QY 538 SRHATYSWRHKNV 551

RESULT 11
ID W64221 standard; Protein: 974 AA.
AC W64221;
DE 06-OCT-1998 (first entry)
DE Human secreted protein from clone CJ145_1.
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory.
OS Homo sapiens.
PN W09827205-A2.
PD 25-JUN-1998.
PF 17-DEC-1997; U23330.
PR 16-DEC-1997; US-991872.
PR 18-DEC-1996; US-769192.
PR 13-JAN-1997; US-783401.
PA (GEM) GENERICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-362774/31.
DR N-P5DB: V44295.
DR New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
PS Claim 17j; Page 71-74; 110pp; English.
CC This sequence represents a novel secreted protein from clone CJ145_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
SQ Sequence 974 AA;

Query Match 9.5%; Score 390; DB 33; Length 974;
Best Local Similarity 29.8%; Pred. No. 6.69e-25;
Matches 136; Conservative 95; Mismatches 174; Indels 52; Gaps 36;

Db 113 kdechfnfklvllknddalfvcgtnafnpscnykmdtlepfdfgdfsgmarpcydakhan 172
QY 91 KRDCENYIT-LLERRSEGLACGTNARHPSCNVLNVTGTVPLG-EMRGVA--PFSPDENS 146
Db 173 valfadgklysatvt-dflaidaviyrs!-gesp--tlrtvkhdkwkkepfvgavdqv 228
QY 147 LVLFEGDEVISTIRKQEYNG-KIPFRIRGESELYTSDTVNQNFQIKATLVHQDQAYD 205
Db 229 dylyfffreiaveyntmgkvfprvaqvcnkdmggsqrvlekqwtfsfkarlncsvpgds 288
QY 206 DKIIYFFREDPNKNPEAPLNVSRAOLCRGQGGESSLSVSKWNTFLKAMLVCSDAATN 265
Db 289 hfyfnllqavtdviring--rdv-vlatfstpynsipsavcaydmldiasvftorfkeq 345
QY 266 K-NFNRLQDVFLLPDPGQWRDRTRYGVFSNPNW-Y--SACVCYSLGDDIKV--RTSSL 319
Db 346 kspdstwtpvderypkprpcagsssleryataneftpddtlnfikhplmdavpsif 405
QY 320 KGYHSS---LPNPRPGKCLPD--QQPIPTETQVADRHPDVA-Q--RVEP-M--G-P-L- 365

Mon Jul 12 06:29:34 1999

US-09-041-236-2-45.rag

Page 9

OY 471 CLMSRDPYCGWDQGR-CISIYSSE-RSVLQSIINPAE-PHKECPN 511

Search completed: Thu Jul 8 20:06:33 1999
Job time : 34 secs.

THIS PAGE BLANK (USPTO)

WIREH

(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 8 20:05:05 1999; MasPar time 23.91 Seconds

Tabular output not generated. 923.550 Million cell updates/sec

Title: >US-09-041-236-2
Description: (55-605) from US09041236.pap (45 of 45)
Perfect Score: 4122
Sequence: 1 VWVGGRKVLDFPFGKNA.....GSYFRAQHWQLLPEDGIMA 551

Scoring table: PAM 150

Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: pir60

1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 48.866; Variance 86.191; scale 0.567

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	1759	42.7	653	2	T03102 semaphorin homolog A3	0.00e+00
2	737	17.9	748	2	I48744 semaphorin A - mouse	1.08e-134
3	725	17.6	751	2	I48748 semaphorin E - mouse	6.27e-132
4	716	17.4	749	2	G01856 semaphorin V - human	7.40e-130
5	688	16.7	753	2	G02173 semaphorin III family	2.00e-123
6	681	16.5	772	2	A49069 collapsin - chicken	8.07e-122
7	667	16.2	666	2	I59169 semaphorin III - mouse	1.30e-118
8	666	16.2	772	2	I48747 semaphorin D - mouse	1.20e-118
9	661	16.0	771	2	D49423 semaphorin III precursor	3.07e-117
10	618	15.0	782	2	I48746 semaphorin C - mouse	1.98e-107
11	522	12.7	834	2	S65498 M-sema F protein precursor	9.60e-86
12	480	11.6	760	2	I48745 semaphorin B - mouse	2.8e-76
13	470	11.4	403	2	A42521 A39R protein - vaccin	3.75e-74
14	463	11.2	441	2	S29921 hypothetical protein	1.33e-72
15	448	10.9	1074	2	JC5928 semaphorin F precursor	2.74e-69
16	447	10.8	730	2	JH0798 fasciclin IV precursor	4.54e-69
17	379	8.2	711	2	A49423 semaphorin I precursor	3.13e-54
18	367	8.9	656	2	B49423 semaphorin I - fruit	1.20e-51
19	362	8.8	724	2	C49423 semaphorin II precursor	1.42e-50
20	293	7.1	295	2	JQ1775 sal19R protein - vacc	5.06e-36
21	196	4.8	1884	2	JC4976 plexin 2 precursor -	9.27e-17
22	170	4.1	1872	2	JC4976 plexin 3 precursor -	5.49e-12
23	168	4.1	1894	2	JC4980 plexin 1 precursor -	1.25e-11

24	171	4.1	1905	2	I51553	Plexin - African claw	3.63e-12
25	157	3.8	122	2	J01845	14R protein - variola	1.07e-09
26	157	3.8	122	2	H36852	A43R protein - variol	1.07e-09
27	155	3.8	142	2	JQ1776	Sal19R protein - vacc	2.38e-09
28	112	2.7	775	2	E70320	polyribonucleotide nu	1.87e-02
29	105	2.5	227	2	S09922	hypothetical protein	1.84e-01
30	105	2.5	235	1	Q0BEC9	HXLf4 protein - human	1.84e-01
31	102	2.5	275	2	A34866	T-cell surface protei	4.76e-01
32	103	2.5	406	2	B89064	conserved hypothetical	3.48e-01
33	99	2.4	275	2	S08464	T-cell alloantigen RT	1.20e+00
34	99	2.4	460	2	S35772	translation elongatio	1.20e+00
35	99	2.4	563	2	S78224	virulence-associated	1.20e+00
36	99	2.4	591	2	S26565	virulence-associated	1.20e+00
37	99	2.4	591	2	S09498	virulence-associated	1.20e+00
38	99	2.4	591	2	S26664	virulence-associated	1.20e+00
39	99	2.4	593	2	S15215	virulence-associated	1.20e+00
40	99	2.4	597	2	A46050	thyroid/steroid recep	1.20e+00
41	97	2.4	630	2	T00352	hypothetical protein	2.21e+00
42	98	2.4	970	2	C57282	ankyrin-related prote	1.63e+00
43	98	2.4	979	2	B57282	ankyrin-related prote	1.63e+00
44	99	2.4	1375	2	JC5148	hepatocyte growth fac	1.20e+00
45	98	2.4	1786	2	A57282	ankyrin-related prote	1.63e+00

ALIGNMENTS

RESULT	1
ENTRY	T03102
TITLE	semaphorin homolog A3 - alcelaphine herpesvirus 1
ORGANISM	#formal_name alcelaphine herpesvirus 1
DATE	24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS	T03102
REFERENCE	Z14840
#authors	Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal	J. Virol. (1997) 71:6517-6525
#title	Primary structure of the alcelaphine herpesvirus 1 genome.
#accession	T03102
#status	preliminary; translated from GB/EMBL/DBJ
#molecule_type	DNA
#residues	1-653
#label	ENS
#cross-references	EMBL:AF005370; NID:g2337967; PID:g2337970
SUMMARY	#length 653 #molecular-weight 73645 #checksum 5501

Query Match 42.7%; Score 1759; DB 2; Length 653;
Best Local Similarity 46.8%; Pred. No. 0.00e+00;
Matches 252; Conservative 95; Mismatches 180; Indels 11; Gaps 11;

Db	109	VYVGNNIYLFDAHSSNASTALINITSTHRLSTSCENFITLLHNQDTGLACGTN	168
Qy	55	VWVGGRKVLDFPFGKNA SVRTNIGSTKGCGLDKRDCENYITLLRSEGLACGTN	114
Db	169	SOKPSCW-LINNLTQFLGPKGLAPSPSSGNLVLFDPNDTYSTINLYKSLGS-HKFR	226
Qy	115	ARHSCNWLWGVTP-LGEMRGAPSPDENSLVFEDEVISTIRK-QYNGKIPRFR	172
Db	227	RIAGQVELYSDTAMHRPQFOATAVHKNESYDDKIYFFFOENSHSDFKQPHVPRVQ	286
Qy	173	RIRGESELYSDTVMQNPQFIKATIVHODQAYDDKIYFFREDNPKNPAPLNVSRVQ	232
Db	287	VCSDDQGESLSVYKWTTFKLAKLACVDYDTGRIYNELQDIFIWQAPENSWEETLI	346
Qy	233	LCRGDQGESLSVSKWNTFLKAMLCVSDAATNKNFNRLQDVFLLPDPGQWRDTRV	292
Db	347	FLSPWNESAVCVFKDIDHVEKTSKLNKYNHKKLPTPEPGCMKNHNPVETQVADRY	406
Qy	293	FSNPWNESAVCVSLGDDIKVFTSSLKGYHSLPNRPGKGLDQDPIPTETQVADRH	352
Db	407	PEVADPVYQKNAMFPIQSQIYITKLVYRVEYG-GVFWATIFVLTITKTIHIYRYE	465
Qy	353	PEVAQRVEPMGLKPTLFHSKYHKQKAVHRMQASHGETFH-VLYLTDRGTIHKVPE	411
Db	466	DSNTTALNILEINFPQKAPQIQLNDNTNLKLYVNSEWSEVPLDCLSYGNDCFSC	525

QY	139	PFSPDENSELVFEGDEVYSTIRKOEYNGK-IPFRIRIRGESELYTS--DTV-MONPOFIK	194
Db	230	AELIPDSEANDKLYFFERSAE-APOSAPVARIICRLNDDGGHCLL-VNKWSTFLK	287
QY	195	ATIVHQDAYDDKIYFFREDNPKNPEAPLVNSRVAQLCRDGGESSLSVSKWTFKL	254
Db	288	ARLVCSVPGEDGTHFDELQDVFV-QOTQDV-RNPVIYAVFTSSGVSFVSGAVCVYSMA	345
QY	255	AMLVCS---DAATNKNFNRLQDVFLLPDPGQWRDTRVGVFSNP-WNY--SAVCVYSLG	308
Db	346	DIRVFNPGFAHKEGPNYQWMPFGSKMPYRPGTCPGTTPSKMSTKDYDPDEVINPMS	405
QY	309	DIDKVFRTS-SLK-G--YH---SS-LPNRPCKLQPDQ-QP-I-PTEFT--QVAD--R-	351
Db	406	HPLMYQAVYLQRRPLVVRT-GAPYRLTTTIAVDQVDSADG-RYEVLFLGTRDGTQVKIV	463
QY	352	HPEVAQRVEPWG--PLKTPLEHSHYHVKQVAVHRMQASHGETHVLVLTTRDGTIHKVVE	409
Db	464	LPKDDQMEELMEEVEFKDPAVPKVTMTISSRQQLYVASAVGVTHLSLHRQOAYGAAC	523
QY	410	-PGEQHSFAFNIMEIQFRRAAAIQTMSLDAERKLYVSSQWESQVPLDLCEVYGGC	468
Db	524	ADCLLARDPYCANDQACSRYTASSRRSRDQVRHGNPIROCRGFSNANKNAVESQY	583
QY	469	HGCLMSRDPYCGMDQGRCLIS-IYSSER-SVLQSNPAEPHKECP--NPKPKAPLQKVS	524
Db	584	GVAGSAFLQCPSPQATVKW	605
QY	525	A-PNSRYILSCPMSRHATYSW	545
RESULT	6		
ENTRY		A49069	#type complete
TITLE		collapsin - chicken	
ORGANISM		#formal_name Gallus gallus	#common_name chicken
DATE		07-Apr-1994	#sequence_revision 07-Apr-1994
ACCESSIONS		A49069	
REFERENCE		Luo, Y.; Raible, D.; Raper, J.A.	
#authors		Cell (1993) 75:217-227	
#journal		Collapsin: a protein in brain that induces the collapse and	
#accession		paralysis of neuronal growth cones.	
##status		preliminary; not compared with conceptual translation	
##molecule_type		mRNA	
##residues		1-772	#label LUO
##cross-references		GB:U02528; NID:g410078; PID:g410079	
CLASSIFICATION		#superfamily semaphorin	
SUMMARY		#length 772	#molecular-weight 88867
Query Match		16.5%;	Score 681; DB 2; Length 772;
Best Local Similarity		31.5%;	Pred. No. 8.07e-122;
Matches		146;	Conservative 119; Mismatches 150; Indels 48; Gaps 38
Db	166	RGKSPYDPKLLTASLLVDGELYSGTA-ADEFMRDFAIRTLGHHPHRTQHSRLWNDP	224
QY	135	RGVAFSPDENSELVFEGDEVYSTIRKOEYNGK-IPFRIRIRGESELYTS--DIV-MQNP	190
Db	225	RFTSAHLIPESDNPEDDKIYFFRENDAIDGHTGKATHARIGQICKNDFGHRSL-VNWK	283
QY	191	QFIKATIVHQ--DQAYDDKIYFFREDNPKNPEAPLVNSRVAQLCRDGGESSLSVSKW	249
Db	284	TTELKARLICSVPQCNIDHFDLQDVFILM-NSKDP-KNPIYGVFTTSSNIFKGSVAVC	341
QY	250	NTELKALVCS--DA-ATNKNFNRLQDVFLLPDPGQWRDTRVGVFSNPW-N-Y--SAVC	303
Db	342	MYSTMDVRRVFLGPAHRDGPNTQWQVYQGRVYPRPGTCPSKTFGGFDSTKDLDPDEV	401
QY	304	VYSLGDDIKVFRYS-SLK-G--YH--S--LPNRPCKLQPDQ-QP-I-PTEFT--QV-A	349
Db	402	FARSHPANYVPFPIINS-RPIMIKTDVDTQYTOIVVDVDAEDGQ-YDVMFIGTIDTIVL	459

Qy	350	-DR-HPEVAORVPMGDLKPTLPFHSK--YHYQKVAVRHMQASGHGTFHYLYLTDRGTH	405
Db	460	KVWSIPKETWHELEVELEBENTVFREPTVISAMKISTKQOOLYIGSATGVSQPLHRCDV	519
Qy	406	KVVE-PGEQHSFAFNIME-IQPFRAAAIQTWSLDAERKLYVSSQWEVSQVPLDCEV	463
Db	520	YGKACACCLAROPYCAWDGSSCSRFPPTAKRTRRQDIRNGDPLHCSDLQHDHNP	579
Qy	464	YGGCGHCLMSRDPYCGWDGRCISYSS-RSVL-QSINPAEPHKECPN-PPDKAP--	518
Db	580	TLEBKIIYGVNSTFLECSPKSORAIVY-WQFQKQNDHDKVE	621
Qy	519	-LQ-KVSLA-PNSRYILSC-PMESRHATYSWR-HKENVEQSC	556
RESULT	7		
ENTRY	I58169	#type fragment	
TITLE	semaphorin III - mouse (fragment)		
ORGANISM	#formal_name Mus musculus #common_name house mouse		
DATE	26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 04-Sep-1998		
ACCESSIONS	I58169		
REFERENCE	I58169		
#authors	Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.; Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L. Neuron (1995) 14:949-959		
#journal	Semaphorin III can function as a selective chemorepellent to pattern sensory projections in the spinal cord.		
#title	pattern sensory projections in the spinal cord.		
#cross-references	MUID:95267432		
#accession	I58169		
#status	preliminary; translated from GB/EMBL/DBJ		
#molecule_type	mRNA		
#residues	1-666 ##label RES		
#cross-references	GB:L40484; NID:g703189; PID:g703190		
GENETICS			
#gene	SemaIII		
CLASSIFICATION	#superfamily semaphorin		
SUMMARY	#length 666 #checksum 9654		
Query Match	16.28; Score 667; DB 2: Length 666;		
Best Local Similarity	30.18; Pred. No. 1.30e-118;		
Matches	150. Conservative 130; Mismatches 168; Indels 50; Gaps 39;		
Db	60	RKSPYDKLLTASLTIDGELYSGTA-ANFMGRDFAIFRTLGHHPHPIRTEQHDSRLNDP	118
Qy	135	RGVAFSPDENSLSLVFEGDEVISTIRKQYNGK-IPFRRIIRGESELYTS--DTV-MQNP	190
Db	119	RFTSAHLIPSDNPDDKYFFRENAIDGESHGKATHAIGQIKNDGEGHRS-LVNW	177
Qy	191	QFIKATIVHO--DQAYDDKIYFFREDNPKNPEAPLVNSRVAQLCRGDQGGESSLSVKW	249
Db	178	TTFLKARLICSVPGNCIDHFELODVLM-NSKDP-KNPITYGVFTTSSNIFKGSVC	235
Qy	250	NTEFLKALVCS--DA-ATNKNFRLQDVELLPDPSQWRDTRIVYGVFSNPWN-Y--SVC	303
Db	236	MYSMSDVRVRLGPIYAHRDGPNQWVPYPRGTCPSKTEGGFSTKDLDPDVT	295
Qy	304	VYSLGDDIKV---F--RTSS-LK--GYHSSLNPRGKCLPDQ-QPI-PTETF--QV-A	349
Db	296	FARSHPAMYNPVFPINN-REPMIKTDVNYQFTQIVVDVDAEDGQ-YDVMFTGTDVGT	353
Qy	350	-DR-HPEVAORVPMGDLKPTLPFHS--KYHYQKVAVRHMQASGHGTFHYLYLTDRGTH	405
Db	354	KVSVVPKETWHDLEEVLEBMTVFREPTISAMELSTKQOOLYIGSTAGVAQLPLHRC	413
Qy	406	KVVE-PGEQHSFAFNIME-IQPFRAAAIQTWSLDAERKLYVSSQWEVSQVPLDCEV	463
Db	414	YGKACACCLAROPYCAWDGSSCSRFPPTAKRTRRQDIRNGDPLHCSDLQHDHNP	473
Qy	464	YGGCGHCLMSRDPYCGWDGRCISYSS-RSVL-QSINPAEPHKECPN-PPDK--AP	518
Db	474	SLEERIIYGVNSTFLECSPKSORALVYVWQFORNEDRKEEIKMGDHIIRTEOGLLR	533

Qy	519	-LQ-KVSLA-PNSRYLSC-PMESRHATY-SW-RHKENVQSCPEG-HQ-SPNCILFIEN	570
Db	534	LQKDSGNLYLCHAVERGF	551
Qy	571	LTAQOYGHYCEAQEGSY	588
RESULT	8		
ENTRY		I48747	#type complete
TITLE		semaphorin D - mouse	
ORGANISM		#formal_name Mus musculus	#common_name house mouse
DATE		02-Jul-1996	#sequence_revision 02-Jul-1996
		04-Sep-1998	#text_change
ACCESSIONS		I48747	
REFERENCE		I48744	
#authors		Puschel, A.W.; Adams, R.H.; Betz, H.	
#journal		Neuron (1995) 14:941-948	
#title		Murine semaphorin D/collapsin is a member of a diverse gene family and creates domains inhibitory for axonal extension	
		#cross-references MUID:95267431	
		#accession	I48747
		#status	preliminary; translated from GB/EMBL/DDBJ
		#molecule_type	mRNA
		#residues	1-772
		#cross-references	EMBL:X85993; NID:98543329; PID:98543330
GENETICS			
#gene		semD	
CLASSIFICATION		#superfamily	semaphorin
SUMMARY		#length	772
		#molecular-weight	88710
		#checksum	1775
		Query Match	16.2%
		Best Local Similarity	31.0%
		Matches	140; Conservative 119; Mismatches 146; Indels 47; Gaps 38;
Db	166	RGKSPYDPKLLTASLLDGLXSGTA-ADFMGRDFAIFRTLGDHHPRTIQHDSRLNDP	224
Qy	135	RGVAPSPDENSLVLFEGDEVSTIRKQYNGK-IPRFRIRGESELYTS--DTV-MQNP	190
Db	225	RFISAHILIPESNDPEDKVVFFRENAIGEGSHGKATHARIQICKNDFGGHRSI-VNKK	283
Qy	191	QFIKATIVHQ-DQAYDDIKYVFFREDNPDKNPEAPLNSRVAQLCRGDOGGESSLSVKW	249
Db	284	TTFLKARLCSVPGPNGIDHDELODFLM-NSKDP-KNPVYGVFTSSNIFKGSVAC	341
Qy	250	NTELKAWLVCS--DA-ATKNFNRLQDVELPDPSCQWRDTRVYGVFSNPWN-Y--SAVC	303
Db	342	MYSMDSVRRVFLGYPYAHRGDPNYQWVYQGRVYPRPGTSPKTEGGFDSTKDLDPDVIT	401
Qy	304	VYSLGDDIKVFRYS-SLK-G-YH--S--S-LPNRPCKCLPDQ-QPI-PTETF--QV-A	349
Db	402	FGSRHPAMPNPVPPINN-PRMIKTDVNTQFTQIVVDVDAEDGQ-YDVMFISGTVGTVL	459
Qy	350	-DR-HPEAQRVPEMPGLKPTPLFS--KYHQKVAVHRMQASHGETFHVLYLTTRDGTIH	405
Db	460	KVYSVPKETHDLEEVLLKEMTVEREPTTISAMELSTKQOQLYIGTAGVAQLPLRHCDI	519
Qy	406	KVYE-FGEQHSFAFIME-IGPFRRAAIIQWMSDAERKRLVYSSQEVSGVLDLCEV	463
Db	520	YGKACAECCCLARDPYCAWDGSGSCRYFFPAKRRTRRQDIRNGDPLTHGSDLEDHNNHGP	579
Qy	464	YGGGCHGLMSRDPYCGWDGRCISIIYSSE-RSVL-QSINPAEPHKECPN-PKPKD--AP	518
Db	580	SLERIIVGVENSSTLECPSPQSQRALVY-WQ	610
Qy	519	-LQ-KVSLA-PNSRYLSC-PMESRHATYSW-R	546

```

9
RESULT
ENTRY
TITLE
ORGANISM
DATE
D49423      #type complete
semaphorin III precursor - human
#formal_name Homo sapiens #common_name man
06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
04-Sep-1998

```


THIS PAGE BLANK (USPTO)

MPERCH_PP protein - protein database search, using Smith-Waterman algorithm

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Thu Jul 8 20:03:05 1999; MasPar time 16.56 Seconds
Tabular output not generated. 940.780 Million cell updates/sec

Title: >US-09-041-236-2
Description: (55-605) from US09041236.pap (45 of 45)
Perfect Score: 4122
Sequence: 1 VVVGGRGVYLFDPCKNA.....GSYFRAQHWLLPDDGIMA 551

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss:prot37
1:swissprot

Statistics: Mean 49.935; Variance 75.474; scale 0.662

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	470	11.4	403	1	VA39_VACCC	1.88e-86
2	463	11.2	441	1	VA39_VACCV	1.20e-84
3	163	4.0	1871	1	SEX_HUMAN	7.17e-13
4	105	2.5	227	1	US08_HCMVA	3.24e-02
5	102	2.5	275	1	NRT2_RAT	9.54e-02
6	99	2.4	275	1	NRT1_RAT	2.74e-01
7	99	2.4	460	1	EF1A_TRIRE	2.74e-01
8	99	2.4	591	1	VRP2_SALTY	2.74e-01
9	99	2.4	591	1	VRP2_SALCH	2.74e-01
10	99	2.4	591	1	VRP2_SALEN	2.74e-01
11	99	2.4	591	1	VRP2_SALDU	2.74e-01
12	99	2.4	916	1	PMAL_AJCCA	2.74e-01
13	98	2.4	1132	1	DNBI_HSV6U	2.74e-01
14	95	2.3	255	1	HA21_HUMAN	3.87e-01
15	94	2.3	329	1	RBSR_ECOLI	1.07e+00
16	93	2.3	380	1	DP3B_MYCPN	1.49e+00
17	95	2.3	518	1	VL2_HPV5B	2.07e+00
18	95	2.3	518	1	VL2_HPV05	1.07e+00
19	95	2.3	683	1	AMOH_ARTGO	1.07e+00
20	94	2.3	947	1	PM2_YEAST	1.49e+00
21	94	2.3	982	1	POL_HTLV2	1.49e+00
22	93	2.3	1029	1	END1_YEAST	2.07e+00
23	94	2.3	1069	1	ENTK_MOUSE	1.49e+00

24	94	2.3	1115	1	IREL_YEAST	SERINE/THREONINE-PROTE	1.49e+00
25	94	2.3	1200	1	DDX8_CAEEL	PUTATIVE PRE-MRNA SPLI	1.49e+00
26	95	2.3	1390	1	MET_HUMAN	HEPATOCYTE GROWTH FACT	1.07e+00
27	90	2.2	119	1	RL19_MYCPN	50S RIBOSOMAL PROTEIN	5.45e+00
28	90	2.2	213	1	AMEX_BOVIN	AMELOGENIN, CLASS I PR	5.45e+00
29	90	2.2	254	1	HA22_HUMAN	HLA CLASS II HISTOCOMP	5.45e+00
30	91	2.2	348	1	DDL_ENTFA	D-ALANINE-D-ALANINE L	3.96e+00
31	91	2.2	356	1	VP39_NPVLD	MAJOR CAPSID PROTEIN	3.96e+00
32	89	2.2	404	1	NIFS_ECOLI	NIFS PROTEIN HOMOLOG	7.46e+00
33	89	2.2	437	1	RFBB_MYAXA	O-ANTIGEN EXPORT SYSTE	7.46e+00
34	92	2.2	442	1	CEMI_YEAST	3-OXOACYL-(ACYL-CARRIE	2.87e+00
35	92	2.2	460	1	EF1A_NEUCR	ELONGATION FACTOR 1-AL	2.87e+00
36	91	2.2	498	1	ELAS_PSEAE	PSEUDOLYSIN PRECURSOR	3.96e+00
37	91	2.2	598	1	NOT_HUMAN	IMMEDIATE-EARLY RESPON	3.96e+00
38	89	2.2	598	1	RNR1_RAT	REGENERATING LIVER NUC	7.46e+00
39	89	2.2	649	1	RA32_SCHPO	DNA REPAIR PROTEIN RAD	7.46e+00
40	91	2.2	707	1	GNVK_HCMVA	GANCICLOVIR KINASE (EC	3.96e+00
41	92	2.2	1132	1	DNBI_HSV62	MAJOR DNA-BINDING PROT	2.87e+00
42	92	2.2	1302	1	MDR5_DROME	MULTIDRUG RESISTANCE P	2.87e+00
43	92	2.2	1663	1	CO3_MOUSE	COMPLEMENT C3 PRECURSO	2.87e+00
44	91	2.2	1666	1	CO3_CAYPO	COMPLEMENT C3 PRECURSO	3.96e+00
45	90	2.2	1874	1	POLR_KYVWJ	RNA REPLICASE POLYPROT	5.45e+00

ALIGNMENTS

RESULT 1
ID VA39_VACCC STANDARD; PRT; 403 AA.
AC P21062;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE PROTEIN A39.
GN A39R.
OS VACCINIA VIRUS (STRAIN COPENHAGEN).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91021027.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RT "The complete DNA sequence of vaccinia virus."
RL VIROLOGY 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RL VIROLOGY 179:517-563(1990).
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; M35027; G335517; -
PIR; E42521; E42521.
SEQUENCE 403 AA; 45741 MW; FE0DB9AB CRC32;

Query Match 11.4%; Score 470; DB 1; Length 403;
Best Local Similarity 33.5%; Pred. No. 1.88e-86;
Matches 86; Conservative 59; Mismatches 95; Indels 17; Gaps 16;

Db 77 LVCGTNGNPKWK-IDGSDDPKRGHRGVPQNSKVITISYN-ECVLSINISK-EG-I 132
QY 109 LAGCTNARHPSCNWNLVGTWPLGEMRGVAFSPDENSLVLEGEVSTIRKQYNGKI 168
Db 133 KWRRRFDGCGVLDYADNVIPKDG-LRGAFVNDKCTYD-KVYLFTDTIGSKR--I-VK 107
QY 169 PRFRIRGE-S-ELYTSDTVMQNPQIKATIVHQDQAYDDKIYFFREDNPDKNPEAPLN 226

QY 131 LGE-MRGYAPFSPDENSELVFEGDEVYTIKQYNGK-IPRFRIRGESE-LYTSDTVM 187

Db 252 ES 253

QY 188 QN 189

RESULT 6

ID NRTL_RAT STANDARD; PRT; 275 AA.

AC P17982;

DT 01-NOV-1990 (REL. 16, CREATED)

DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1 PRECURSOR (EC 2.4.2.31) (T-CELL

DE NAD(P)(+)-ARGININE ADP-RIBOSYLTRANSFERASE 1) (T-CELL MONO(ADP-

DE RIBOSYL)TRANSFERASE 1) (ALLOANTIGEN RT6.1) (T-CELL SURFACE PROTEIN

DE RT6.1).

GN ART2A OR RT6-A.

OS RATTUS NORVEGICUS (RAT).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.

[1]

RN SEQUENCE FROM N.A.

RC STRAIN-LEWIS A.

RX MEDLINE; 90192088.

RA HAAG F., KOCH F., THIELE H.-G.;

RT "Nucleotide and deduced amino acid sequence of the rat T-cell

RT alloantigen RT6.1";

RL NUCLEIC ACIDS RES. 18:1047-1047(1990).

[2]

RN MUTAGENESIS OF GLN-207.

RX MEDLINE; 96275529.

RA MAEHAMA T., HOSHINO S.-I., KATADA T.;

RT "Increase in ADP-ribosyltransferase activity of rat T lymphocyte

RT alloantigen RT6.1 by a single amino acid mutation.";

RL FEBS LETT. 388:189-191(1996).

CC -!- FUNCTION: HAS NAD+ GLYCOHYDROLASE ACTIVITY AND EXTREMELY LOW ADP-

CC -!- RIBOSYLTRANSFERASE ACTIVITY.

CC -!- CATALYTIC ACTIVITY: NAD(+) + L-ARGININE = NICOTINAMIDE +

CC N2 (ADP-D-RIBOSYL)-L-ARGININE.

CC -!- CATALYTIC ACTIVITY: NAD(+) + H(2)O = NICOTINAMIDE + ADP-RIBOSE.

CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.

CC -!- TISSUE SPECIFICITY: POSTTHYMIC T CELLS.

CC -!- PTM: TWO DISULFIDE BONDS ARE PRESENT (PROBABLE).

CC -!- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE

CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to license@isb-sib.ch.

CC EMBL; X52082; G57168; -.

CC EMBL; M31138; G206804; -.

CC PIR; S08464; S08464.

CC PROSITE; PS01291; ART; 1.

CC PFAM; PF01129; ART; 1.

CC TRANSFERASE: GLYCOSYLTRANSFERASE; GLYCOPROTEIN; NAD; SIGNAL;

CC T-CELL DIFFERENTIATION; GPI-ANCHOR.

FT SIGNAL 1 20

FT CHAIN 21 246 T-CELL ECTO-ADP-RIBOSYLTRANSFERASE 1.

FT PROPER 247 275 HYDROPHOBIC. REMOVED DURING MATURATION

FT LIPID 246 246 (BY SIMILARITY).

FT ACT_SITE 209 209 GPI-ANCHOR (BY SIMILARITY).

FT CARBOHYD 58 58 BY SIMILARITY.

FT MUTAGEN 207 207 POTENTIAL.

FT O->E: INCREASED ADP-RIBOSYLTRANSFERASE

FT ACTIVITY.

FT SEQUENCE 275 AA; 31388 MW; 523B1A84 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 275;

Best Local Similarity 30.6%; Pred. No. 2.74e-01;

Matches 19; Conservative 17; Mismatches 21; Indels 5; Gaps 5;

Db 194 LGVYIKERS-FYPDQEE-VLIPGYEVYQVKTQGYNEIFLDSPKRKKSNYNCLYSSAGTR 251

QY 131 LGE-MRGYAPFSPDENSELVFEGDEVYTIKQYNGK-IPRFRIRGESE-LYTSDTVM 187

Db 252 ES 253

QY 188 QN 189

RESULT 7

ID EFIA_TIRE STANDARD; PRT; 460 AA.

AC P34825;

DT 01-FEB-1994 (REL. 28, CREATED)

DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).

GN TEFL.

OS TRICHODERMA REESEI (HYPOCREA JECORINA).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;

OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.

[1]

RN SEQUENCE FROM N.A.

RC STRAIN-QM9414 / RUTC-30;

RX MEDLINE; 94124021.

RA NAKARI T., ALATALO E., PENTTILA M.;

RT "Isolation of Trichoderma reesei genes highly expressed on glucose-

RT containing media: characterization of the tefl gene encoding

RT translation elongation factor 1 alpha.";

RL GENE 136:313-318(1993).

CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF

CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN

CC BIOSYNTHESIS.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.

CC EF-TU/EF-1A SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to license@isb-sib.ch.

CC EMBL; Z23012; G312887; -.

CC PIR; S35772; S35772.

CC PROSITE; PS00301; EFATOR_GTP; 1.

CC PFAM; PF00009; GTP_EFTU; 1.

CC HSP; P07157; 1AIP.

CC ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; GTP-BINDING.

FT NP_BIND 15 22 GTP (BY SIMILARITY).

FT NP_BIND 92 96 GTP (BY SIMILARITY).

FT NP_BIND 154 157 GTP (BY SIMILARITY).

CC SEQUENCE 460 AA; 49830 MW; B9ABAB2 CRC32;

Query Match 2.4%; Score 99; DB 1; Length 460;

Best Local Similarity 33.3%; Pred. No. 2.74e-01;

Matches 15; Conservative 13; Mismatches 15; Indels 2; Gaps 2;

Db 210 YKGWEKETRAGFTG-KTLEAIDSTIEPKR-PTDKPLRLPLQDV 252

QY 478 YGWDGRCISYSSERSVLSINPAEPKPCPKAPLOKV 522

RESULT 8

ID VRP2_SALTY STANDARD; PRT; 591 AA.

AC P21454;

DT 01-MAY-1991 (REL. 18, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

```
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN MKAA.
OS SALMONELLA TYPHIMURIUM.
OG PLASMID 96 KB VIRULENCE PEX102.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC SALMONELLA.
CC [1]
CC SEQUENCE FROM N.A.
RC STRAIN-TML R56;
RX MEDLINE; 90136009.
RA TAJRA S., RHEN M.;
RT "Identification and genetic analysis of mkaa -- a gene of the
RT Salmonella typhimurium virulence plasmid necessary for intracellular
RT growth."
RL MICROB. PATHOG. 7:165-173(1989).
RN [2]
RN SEQUENCE OF 1-10.
RX MEDLINE; 91244158.
RA TAJRA S., BAUMANN M., RIIKONEN P., SUKUPOLVI S., RHEN M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium."
RL FEMS MICROBIOL. LETT. 61:319-323(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52035; G46898; -.
CC PIR; S09498; S09498.
CC PLASMID; VIRULENCE.
CC DOMAIN 367 373 POLY-PRO.
CC SEQUENCE 591 AA; 65341 MW; BE4A439A CRC32;
CC -----
Query Match 2.4%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 2.74e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLGKTAARLSDPQAASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RDCENYITLLERRSEGLLACGTNARHPSCWNLVNGTVVPLGEMRGYAPFSPDENSILVLF 151
Db 220 GNEAGDRSAMRYSKYV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPFRFRIGSESELY 181
CC -----
RESULT 10
ID VRP2_SALCH STANDARD; PRT; 591 AA.
AC P17450;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN (PROTEIN M2, IN MBA REGION).
OS SALMONELLA CHOLERAE-SUIS (SALMONELLA ENTERICA).
OG PLASMID PKDSC50.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC SALMONELLA.
CC [1]
CC SEQUENCE FROM N.A.
RC STRAIN-RF-1;
RX MEDLINE; 90245675.
RA MATSUI H.;
RT "Nucleotide sequences of genes encoding 32 kDa and 70 kDa
RT polypeptides in mba region of the virulence plasmid, PKDSC50, of
Salmonella choleraesuis."
RL NUCLEIC ACIDS RES. 18:2181-2181(1990).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52035; G46898; -.
CC PIR; S09498; S09498.
CC PLASMID; VIRULENCE.
CC DOMAIN 367 373 POLY-PRO.
CC SEQUENCE 591 AA; 65341 MW; BE4A439A CRC32;
CC -----
Query Match 2.4%; Score 99; DB 1; Length 591;
Best Local Similarity 23.3%; Pred. No. 2.74e-01;
Matches 21; Conservative 25; Mismatches 41; Indels 3; Gaps 3;
Db 162 HDSNGILHLGKTAARLSDPQAASHTAQW-LVEESVTPAGE-HIYYSYLAENGNDVDLN 219
QY 92 RDCENYITLLERRSEGLLACGTNARHPSCWNLVNGTVVPLGEMRGYAPFSPDENSILVLF 151
Db 220 GNEAGDRSAMRYSKYV-QYGNATPAADLY 248
QY 152 GDEVYSTIRKOEYNGKIPFRFRIGSESELY 181
CC -----
RESULT 10
ID VRP2_SALEN STANDARD; PRT; 591 AA.
AC P55220;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 65 KD VIRULENCE PROTEIN.
GN SPVB.
OS SALMONELLA ENTERITIDIS.
OG PLASMID PNL2001.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
CC SALMONELLA.
CC [1]
CC SEQUENCE FROM N.A.
RC STRAIN-AL1190;
RX MEDLINE; 94362897.
RA SUZUKI S., KOMASE K., MATSUI H., ABE A., KAWAHARA K., TAMURA Y.,
RA KIJIMA M., DANBARA H., NAKAMURA M., SATO S.;
RT "Virulence region of plasmid PNL2001 of Salmonella enteritidis."
RL MICROBIOLOGY 140:1307-1318(1994).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D14490; G517164; -.
CC PLASMID; VIRULENCE.
CC DOMAIN 367 373 POLY-PRO.
CC SEQUENCE 591 AA; 65350 MW; 9C5B2EE3 CRC32;
CC -----
Query Match 2.4%; Score 99; DB 1; Length 591;
```


CC SEQUENCE FROM N.A.
RX MEDLINE: 95266321.
RA COMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution."
RL VIROLOGY 209:29-51(1995).
CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES' DNA-BINDING PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X92436; GI044871; .
DR EMBL: X83413; G854020; .
DR PFAM: PF00747; Viral_DNA_bp; 1.
KW DNA-BINDING; DNA REPLICATION; ZINC-FINGER; NUCLEAR PROTEIN.
FT ZN_FING 459 475 C4-TYPE.
SQ SEQUENCE 1132 AA; 127762 MW; DB699A02 CRC32;

Query Match 2.4%; Score 98; DB 1; Length 1132;
Best Local Similarity 28.2%; Pred. No. 3.87e-01;
Matches 24; Conservative 18; Mismatches 37; Indels 6; Gaps 6;

Db 435 WNLNRMVYNAGNAHTETI-YNHLVNCNANLCEFCGDKC-C-QSCIGTAMVRVGTRLPAIP 491
QY 297 WNSAVCVISLGDIDKVFRTSSLGKGYHSSLPNRPGRKCLPDQDQPIPTETFOVADRHEVA 356

Db 492 KNRK-KEPLVMSMF-SRY-YAEVDI 513
QY 357 QRVEPMGLKTPLESHKYHQKQAV 381

RESULT 14
ID HA21_HUMAN STANDARD; PRT; 255 AA.
AC P01908;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) ALPHA CHAIN PRECURSOR
DE (DC-4 ALPHA CHAIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84168117.
RA AUFFRAY C., LILLIE J.W., ARNOT D., GROSSBERGER D., KAPPES D.,
RA STROMINGER J.L.;
RT "Isotypic and allelotypic variation of human class II
RT histocompatibility antigen alpha-chain genes."
RL NATURE 308:327-333(1984).
RN [2]
RP SEQUENCE OF 40-255 FROM N.A.
RX MEDLINE: 83065171.
RA AUFFRAY C., KORMAN A.J., ROUX-DOSSETO M., BONO R., STROMINGER J.L.;
RT "cDNA clone for the heavy chain of the human B cell alloantigen DC1:
RT strong sequence homology to the HLA-DR heavy chain."
RL PROC. NATL. ACAD. SCI. U.S.A. 79:6337-6341(1982).
RN [3]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J00199; .; NOT_ANNOTATED_CDS.
DR PIR: A02212; HLHUCD.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; Ig; 1.
DR PFAM: PF00993; MHC_II_alpha; 1.
DR HSP: P01910; IIAK.
KW MHC II; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 23
FT CHAIN 24 255 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT DQ(1) ALPHA CHAIN.
FT DOMAIN 24 110 EXTRACELLULAR ALPHA-1.
FT DOMAIN 111 204 EXTRACELLULAR ALPHA-2.
FT DOMAIN 205 217 CONNECTING PEPTIDE.
FT TRANSMEM 218 240
FT DOMAIN 241 255 CYTOPLASMIC TAIL.
FT DISULFID 133 189 BY SIMILARITY.
FT CARBOHYD 104 104 POTENTIAL.
FT CARBOHYD 144 144 POTENTIAL.
SQ SEQUENCE 255 AA; 28105 MW; 5C46DD59 CRC32;

Query Match 2.3%; Score 95; DB 1; Length 255;
Best Local Similarity 30.9%; Pred. No. 1.07e+00;
Matches 17; Conservative 12; Mismatches 24; Indels 2; Gaps 2;

Db 25 DIVADSVAAQLG-VNLXQSYGPSQYSHEFDGDEFFVYDLERKETVWLPLFRFR 78
QY 122 NLVNGTVPLGMRGYAPSPDENSILVFEDEV-YSTIRKQEVNGKIPRRIR 175

RESULT 15
ID RBSR_ECOLI STANDARD; PRT; 329 AA.
AC P25551;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOSE OPERON REPRESSOR.
GN RBSR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6 AND 250-256.
RC STRAIN=K12;
RX MEDLINE: 93278299.
RA MAUZY C.A., HERMOSON M.A.;
RT "Structural and functional analyses of the repressor, RbsR, of the
RT ribose operon of Escherichia coli."
RL PROTEIN SCI. 1:831-842(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE: 93315143.
RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication."
RL GENOMICS 16:551-561(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE: 93278300.
RA MAUZY C.A., HERMOSON M.A.;
RT "Structural homology between rbs repressor and ribose binding protein
RT implies functional similarity."
RL PROTEIN SCI. 1:843-849(1992).
CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE RIBOSE RBSDACBK
CC OPERON. RBSR BINDS TO A REGION OF PERFECT DYAD SYMMETRY SPANNING

Search completed: Thu Jul 8 20:03:37 1999
Job time : 32 secs.

Result No.	Score	Query δ		Length	DB	ID	Description	Pred. I
		Match	Length					
1	4122	100.0	666	4	075326		SEMAPHORIN L.	0.00e
2	2131	51.7	393	11	088371		SEMAPHORIN L. (FRAGMENT	0.00e
3	1759	42.7	653	14	094906		SIMILAR TO GENBANK ACC	0.00e
4	7737	17.7	748	11	062117		SEMAPHORIN A PRECURSOR	1.37e
5	7285	17.6	751	11	062117		SEMAPHORIN E PRECURSOR	1.90e
6	7116	17.4	749	4	013214		SEMAPHORIN V.	4.31e
7	7116	17.4	750	4	093018		SEMAPHORIN V.	4.31e
8	710	17.2	751	4	099985		SEMAPHORIN E.	1.60e
9	705	17.1	751	13	042236		COLLAPSN 3.	3.24e
10	688	16.7	753	4	013372		SEMAPHORIN III FAMILY	8.91e
11	682	16.5	754	11	088633		SEMAPHORIN IV ISOFORM	3.27e
12	681	16.5	772	13	090607		COLLAPSN.	5.96e
13	682	16.5	785	11	088632		SEMAPHORIN IV ISOFORM	3.27e
14	681	16.5	785	4	013275		SEMAPHORIN IV.	5.96e
15	679	16.5	785	4	015704		SEMAPHORIN.	1.98e
16	668	16.2	772	11	063548		SEMAPHORIN III/COLLAPS	1.45e
17	666	16.2	772	11	062180		SEMAPHORIN D PRECURSOR	4.81e
18	661	16.0	771	4	041563		SEMAPHORIN-III.	9.63e
19	618	15.0	782	11	062179		SEMAPHORIN C (SEM C) (1.38e
20	614	14.9	775	11	070275		SEMAPHORIN H.	1.49e

```
Db 387 VAORVEPMGLKTPLFHSHYHYQKVAVHRMQASHGETFHVLTLDTRGTIHKVVEPGSOE 446
Qy 355 VAORVEPMGLKTPLFHSHYHYQKVAVHRMQASHGETFHVLTLDTRGTIHKVVEPGSOE 414
Db 447 HSFANIMEIQFRAAAATQTSLSAERKLVSSQWESVQVPLDCEVYGGCHGCLMS 506
Qy 415 HSFANIMEIQFRAAAATQTSLSAERKLVSSQWESVQVPLDCEVYGGCHGCLMS 474
Db 507 RDPYCGWDOGRICISYSSRSVLSQINPAEPHKECPNPKPDKAPLQKVSLAPNSRYILSC 566
Qy 475 RDPYCGWDOGRICISYSSRSVLSQINPAEPHKECPNPKPDKAPLQKVSLAPNSRYILSC 534
Db 567 PMESRHATYSWRHKNVEQSCPGHQPNCILFIENLTAQQYGHYFCEAQEGSYFREAQH 626
Qy 535 PMESRHATYSWRHKNVEQSCPGHQPNCILFIENLTAQQYGHYFCEAQEGSYFREAQH 594
Db 627 WQLLPEDGIMA 637
Qy 595 WQLLPEDGIMA 605

RESULT 2
ID 088371 PRELIMINARY; PRT; 393 AA.
AC 088371:
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN L (FRAGMENT).
GN SEMAL.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 98389619.
RA LANGE C., LIEHR T., GOEN M., GEBHART E., FLECKENSTEIN B., ENSER A.;
RT "New eukaryotic semaphorins with close homology to semaphorins of DNA
RL GENOMICS 51:340-350(1998).
DR EMBL; AF030699; G3523117; -.
FT NON-TER 393
SQ SEQUENCE 393 AA; 43909 MW; 7CB8193C CRC32;

Query Match 51.7%; Score 2131; DB 11; Length 393;
Best Local Similarity 92.6%; Pred. No. 0.00e+00;
Matches 287; Conservative 12; Mismatches 10; Indels 1; Gaps 1;

Db 84 VWGGRGKYYHNFPEGKNASVRTNIGSTKGCQDKQDCGNYITLLERRGNGLVCGTN 143
Qy 55 VWGGRGKYLDFPEGKNASVRTNIGSTKGCCLDKDCENYITLLERRSEGLLACGTN 114
Db 144 ARKPCSNLVNSVMSLGEMKGYAPFSPDENSLVLFEGDEVYSTRKQYNGKIPRFR 203
Qy 115 ARHPCSNLVNVTVP-PLGEMRGYAPFSPDENSLVLFEGDEVYSTRKQYNGKIPRFR 173
Db 204 IRGESELYTSDVMQNPQFIKATIVHQDQAYDDKIYFFREDNPKNPEAPLVNSRVAQL 263
Qy 174 IRGESELYTSDVMQNPQFIKATIVHQDQAYDDKIYFFREDNPKNPEAPLVNSRVAQL 233
Db 264 CRGDGGSLSVSKWNTFLKAMLVCSDAATNRNRLQDVFLPDPSQGWDRTRVYGV 323
Qy 234 CRGDGGSLSVSKWNTFLKAMLVCSDAATNRNRLQDVFLPDPSQGWDRTRVYGV 293
Db 324 SNPNYSAVCYSLGDIDRVFTSSLKGYHMGCLNPRGMCLPKKQPIPTETFOVADSH 383
Qy 294 SNPNYSAVCYSLGDIDRVFTSSLKGYHSSLNPRPGKCLPDQOPIPTETFOVADRH 353
Db 384 EVAQVEPMG 393
Qy 354 EVAQVEPMG 363
```

```
RESULT 3
ID Q64906 PRELIMINARY; PRT; 653 AA.
AC Q64906:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GENBANK ACCESSION NUMBER L26081.
OS ALCELAAPHINE HERPESVIRUS 1.
OC VIRUSES; DSONA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
[1]
RN SEQUENCE FROM N.A.
RX STRAIN=C500;
RX MEDLINE; 97201573.
RA ENSER A., FLECKENSTEIN B.;
RT "Alcelaphine herpesvirus type 1 has a semaphorin-like gene.";
RL J. GEN. VIROL. 76:1063-1067(1995).
[2]
RN SEQUENCE FROM N.A.
RX STRAIN=C500;
RX MEDLINE; 97404659.
RA ENSER A., PFLANZ R., FLECKENSTEIN B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. VIROL. 71:6517-6525(1997).
[3]
RN SEQUENCE FROM N.A.
RX STRAIN=C500;
RX MEDLINE; 97404659.
RA ENSER A., PFLANZ R., FLECKENSTEIN B.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U18243; G1000711; -.
DR EMBL; AF005370; G2337970; -.
SQ SEQUENCE 653 AA; 73645 MW; CCCB53C9 CRC32;

Query Match 42.7%; Score 1759; DB 14; Length 653;
Best Local Similarity 46.8%; Pred. No. 0.00e+00;
Matches 252; Conservative 95; Mismatches 180; Indels 11; Gaps 11;

Db 109 YVVGNGNTIYLFDAHSSNASTALINITSTHNTHRLSSTCFITLLHNQTDGLLACGTN 168
Qy 55 VWGGRGKYLDFPEGKNASVRTNIGSTKGCCLDKDCENYITLLERRSEGLLACGTN 114
Db 169 SQKPSCW-LINNLTQFLGPKGLAPSPSSGNLVLEDQNDYTYINLYKSLGS-HKFR 226
Qy 115 ARHPCSNLVNVTVP-LGEMRGYAPFSPDENSLVLFEGDEVYSTRKQYNGKIPRFR 172
Db 227 RIAGQVELYTSDTAMHRPQFVQATAVHKNESYDDKIYFFFOENSHSDPKQFPHTVPRVQ 286
Qy 173 IRGESELYTSDVMQNPQFIKATIVHQDQAYDDKIYFFREDNPKNPEAPLVNSRVAQ 232
Db 287 VCSDDQGSSELSVYKWTTFLLKARLACVDYDTGRIYNELODIFIWOAPENSWEETLYGL 346
Qy 233 LCRGDGGSSELSVSKWNTFLKAMLVCSDAATNRNRLQDVFLPDPSQGWDRTRVYGV 292
Db 347 FLSPWNFSVAVCTVVKDIDHVFKTSKLKYNHKLPTPRPGCMKNHQHVPETFOVADRY 406
Qy 293 FSNPNWYSAVCYSLGDIDRVFTSSLKGYHSSLNPRPGKCLPDQOPIPTETFOVADRH 352
Db 407 PEVADPYQKNAMFPPIQSKYIYTKLVYRVEYG-VGFWATIEVLTIKGTIHYRYE 465
Qy 353 PEVADPYQKNAMFPPIQSKYIYTKLVYRVEYG-VGFWATIEVLTIKGTIHYRYE 411
Db 466 DSNSTTALNILEINPFQKPAPIQNIILLDNTNLKLYVSEWSEVPLDCLSVYNGDGFSC 525
Qy 412 EOEHSFAFNIMEIQFRAAAATQTSLSAERKLVSSQWESVQVPLDCEVYGGCHG 471
Db 526 FMSRDPICLTYWNNTC-S-FK-QRVSVETGPGANRTLSEMGDGHVAPTVKHKQVSIPLSN 582
Qy 472 LMSRDPICLTYWNNTC-S-FK-QRVSVETGPGANRTLSEMGDGHVAPTVKHKQVSIPLSN 529
Db 583 SYLSCPASVNHADYFWTKDGTETKRCVHTKHKNDICILLIANSTTATNGTHVCNKKEDS 640
Qy 530 YYLSCPMEHSRATYSWRHKNVEQSCPGHQPNCILFIENLTAQQYGHYFCEAQEGS 587
```



```
QY 469 HGCLMSRDPYCGWDQGRCSIIYSSE--RSVLQINPAEPHKECP--NPKDPKAPLOKVS 524
Db 582 GVRNNSTELECAPKSPQASIKWLOKDK 609
QY 525 A-PNSRYILSCPMEHSRATYSWR-HKEN 550

RESULT 6
ID Q13214 PRELIMINARY; PRT; 749 AA.
AC Q13214;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96210603.
RA SEKIDO Y., BADER S., LATIF F., CHEN J.Y., DUH F.M., WEI M.H.,
ALBANESI J.P., LEE C.C., LERMAN M.I., MINNA J.D.;
RT "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
cancer deletion region and demonstrate distinct expression
patterns.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4120-4125(1996).
DR EMBL; U28369; G974284; -.
DR PFAM; PF00047; ig; 1.
SQ SEQUENCE 749 AA; 83121 MW; 6138C5D5 CRC32;

Query Match 17.4%; Score 716; DB 4; Length 749;
Best Local Similarity 33.0%; Pred. No. 4.31e-148;
Matches 167; Conservative 115; Mismatches 170; Indels 54; Gaps 41;

Db 112 ECMNFVKLLHAYNRTHLLACGTGAFTCAFEVGHRAEPEVLRDPGRIEDKGKSPYD 171
QY 93 DCENYITLLER-RSEGLACGTNARHPSC-WNLV-N-G--TVVPL--G--EM-RGYAPFS 141
Db 172 PRHRAASVLVGEELYSGVA-ADLMGRDFTIFRSLGORPSLRTEPHDSRWLNPEKFKVFW 230
QY 142 PDENSLVLFEGDEVYSIRKOEYNGK-IPFRIRIGSELYT-S-DTV-MONPOFIKAT- 196
Db 231 IPSENPDDDKIYFFRETAVEAAPALGRLSVRVGQICRNDVGGQSL-VNKKWTFILKA 289
QY 197 IVHODQAYDDKIYFFREDNPKNPE-APLNVSRVAQLCRDQGGESSLSVKWNTFLKA 255
Db 290 RLVCSPGVGEGDTHFDQLQDVFLLS-SRDH-RTPLLYAVFSTSSIFQGSVAVCYSMNDV 347
QY 256 MLVCS--DAATNKNFNRLQDVFLLPDPGQWRDTRVGVFSPWN-Y--SAVCVYSLGDI 310
Db 348 RRAFLGPFANKEGPMHGWYSYQGRVYPRGMCPSKTFGTFSSTKDPDDVIOFARNHPL 407
QY 311 DK--V--F--RTSSL-K--GYHSSLNPRPGKCLPQO-QPIP-TETF--QVAD--R-HPE 354
Db 408 MYSVLPTGG-R-PLFLQVGANYTFTQIAADRVAAADGH-YDVLFTGTDVGTVLKVISVP 464
QY 355 VAQRVEMGPLKTPLF-H--SKYHYQKVAVHRMQASHGETFHVLYLTDRGTHKVVVE-P 410
Db 465 KGSRPSAEGLLLELHVFDSDSAVTSMQISSKRHLQYVARSVAQAIALHRCAAHGRVCT 524
QY 411 -GEOHSAFNAIMEIQFRRAAAIQTMSLDAERKLYVSSQWEVSQVPLDICEVYGGCH 469
Db 525 ECCLARDPYCANDGVACTRFPQSAKRFRQDVNRGDPSTLCSGDSRRPALLEHKKVGV 584
QY 470 GCLMSRDPYCGWDQGRCSIIYSSESVL--QSNINPAEPHKECPNPKDPKAPL-QKVS 526
Db 585 GSSAFLECEPRSLQARVENTFORAGV 610
QY 527 NSRYILSCPMEHSRATYSWR-HKENV 551

RESULT 7
ID Q99985 PRELIMINARY; PRT; 751 AA.
AC Q99985;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E.
```

```
ID Q93018 PRELIMINARY; PRT; 750 AA.
AC Q93018;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN V.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA DANTE M., WAMSLEY P.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U73167; G2880035; -.
DR PFAM; PF00047; ig; 1.
SQ SEQUENCE 750 AA; 83034 MW; CBC87E34 CRC32;

Query Match 17.4%; Score 716; DB 4; Length 750;
Best Local Similarity 33.0%; Pred. No. 4.31e-148;
Matches 167; Conservative 115; Mismatches 170; Indels 54; Gaps 41;

Db 113 ECMNFVKLLHAYNRTHLLACGTGAFTCAFEVGHRAEPEVLRDPGRIEDKGKSPYD 172
QY 93 DCENYITLLER-RSEGLACGTNARHPSC-WNLV-N-G--TVVPL--G--EM-RGYAPFS 141
Db 173 PRHRAASVLVGEELYSGVA-ADLMGRDFTIFRSLGORPSLRTEPHDSRWLNPEKFKVFW 231
QY 142 PDENSLVLFEGDEVYSIRKOEYNGK-IPFRIRIGSELYT-S-DTV-MONPOFIKAT- 196
Db 232 IPSENPDDDKIYFFRETAVEAAPALGRLSVRVGQICRNDVGGQSL-VNKKWTFILKA 290
QY 197 IVHODQAYDDKIYFFREDNPKNPE-APLNVSRVAQLCRDQGGESSLSVKWNTFLKA 255
Db 291 RLVCSPGVGEGDTHFDQLQDVFLLS-SRDH-RTPLLYAVFSTSSIFQGSVAVCYSMNDV 348
QY 256 MLVCS--DAATNKNFNRLQDVFLLPDPGQWRDTRVGVFSPWN-Y--SAVCVYSLGDI 310
Db 349 RRAFLGPFANKEGPMHGWYSYQGRVYPRGMCPSKTFGTFSSTKDPDDVIOFARNHPL 408
QY 311 DK--V--F--RTSSL-K--GYHSSLNPRPGKCLPQO-QPIP-TETF--QVAD--R-HPE 354
Db 409 MYSVLPTGG-R-PLFLQVGANYTFTQIAADRVAAADGH-YDVLFTGTDVGTVLKVISVP 465
QY 355 VAQRVEMGPLKTPLF-H--SKYHYQKVAVHRMQASHGETFHVLYLTDRGTHKVVVE-P 410
Db 466 KGSRPSAEGLLLELHVFDSDSAVTSMQISSKRHLQYVARSVAQAIALHRCAAHGRVCT 525
QY 411 -GEOHSAFNAIMEIQFRRAAAIQTMSLDAERKLYVSSQWEVSQVPLDICEVYGGCH 469
Db 526 ECCLARDPYCANDGVACTRFPQSAKRFRQDVNRGDPSTLCSGDSRRPALLEHKKVGV 585
QY 470 GCLMSRDPYCGWDQGRCSIIYSSESVL--QSNINPAEPHKECPNPKDPKAPL-QKVS 526
Db 586 GSSAFLECEPRSLQARVENTFORAGV 611
QY 527 NSRYILSCPMEHSRATYSWR-HKENV 551

RESULT 8
ID Q99985 PRELIMINARY; PRT; 751 AA.
AC Q99985;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEMAPHORIN E.
```


Db 406 HPLMYQAVYPLQRRPLVVRT-GAPVRLTTIAVDQVDSADG-RYEVLFLGTDRGTQVKVIV 463
Qy 352 HPEVAQRVEPMG--PLKTPLFHSHYQKVAVHRMQASHGETFHVLYLTDRGTIHKVVE 409
Db 464 LPKDDQEMEELMEVEFKDPAPVKTMTISSKQQQLYVASAVGVTLSLHRCQAYGAAC 523
Qy 410 -PGQEHSFAFNIEMIQFRRAAIQTMSLDAERKLYVSSQWESVQVPLDCEVYGGC 468
Db 524 ADCCLARDPYCAMDQACSRVTASSKRRSRQDVHRGNPIRQCGRFNSNANKNAVESVQY 583
Qy 469 HGCLMSRDPYCGWDGRCIS-IYSSER-SVLQSNPAEPHKECP--NPKDPKAPLQKVS 524
Db 584 GVAGSAAFLEQCPSPQATVKW 605
Qy 525 A-PNSRYLLSCPMESRHATYSW 545

RESULT 11
ID O88633 PRELIMINARY; PRT; 754 AA.
AC O88633;
DT 01-NOV-1998 (TREMBLREL, 08, CREATED)
DT 01-NOV-1998 (TREMBLREL, 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL, 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN IV ISOFORM A.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RT ECKHARDT F., MEYERHANS A.;
RA "Molecular cloning and expression pattern of a murine semaphorin
homologous to H-sema IV";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF080091; G377768; -.
SQ SEQUENCE 754 AA; 85133 MW; EC05BDF3 CRC32;

Query Match 16.5%; Score 682; DB 11; Length 754;
Best Local Similarity 31.8%; Pred. No. 3.27e-139;
Matches 141; Conservative 119; Mismatches 140; Indels 43; Gaps 31;

Db 171 PYDPKLDATASALINELVAGVVI-DFMGTDAAIFTLGKQTAMRTDQVNSRWLNDPSPH 229
Qy 139 PFPDENSLVLFEGDEVISTIRKQYNGK-IPFRIRIGSELYTS--DIV-MQNPQIK 194
Db 230 AELIPDSAEKNDKLYFFFRERSAB-APQNPVAVYARIGRLNDGGHCCL-VNKMSTFL 287
Qy 195 ATIVHQD-QAYDDKIYFFREDNPNKPEAPLVNSRVAQLCRGOGGESSLSVKWNTFL 253
Db 288 KARLVCSVPGEDGITHFDELQDVV-QOTQDI-RNPVIYAVFTSSGVSFRGSVAVCYSM 345
Qy 254 KAMLVCS---DAATNKNFNRLQDVFLLPDPGQWRDTRVYGVFSNP-WNY--SAVCVYSL 307
Db 346 ADIRMVENGPAHKEGPNYQWMPFSGKMPYPRPGTFCPGTTPSMKSTKDPYDEVINPMR 405
Qy 308 GOIDKVFRTS-SLK-G-YH-----SS-LPNRPGRKCLPDQ-OP-I-PTETP--QVAD--R 351
Db 406 THPLMYQAVYPLQRRPLVVRT-GAPVRLTTIAVDQVDAADG-RYEVLFLGTDRGTQVKVI 463
Qy 352 -HPEVAQRVEPMG--PLATPLFHSKHYQKVAVHRMQASHGETFHVLYLTDRGTIHKV 408
Db 464 VLPKDDQVEELMLEVEVFKEPAPVKTMTISSKQQQLYVASAVGVTLSLHRCQAYGA 523
Qy 409 E-PGQEHSFAFNIEMIQFRRAAIQTMSLDAERKLYVSSQWESVQVPLDCEVYGG 467
Db 524 CADCCCLARDPYCAMDQACSRVTASSKRRSRQDVHRGNPIRQCGRFNSNANKNAVESVQ 583
Qy 468 CHGCLMSRDPYCGWDGRCIS-IYSSER-SVLQSNPAEPHKECP--NPKDPKAPLQKVS 523
Db 584 YGVAGSAAFLEQCPSPQATVKW 606
Qy 524 LA-PNSRYLLSCPMESRHATYSW 545

RESULT 12
ID Q90607 PRELIMINARY; PRT; 772 AA.
AC Q90607;
DT 01-NOV-1996 (TREMBLREL, 01, CREATED)
DT 01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL, 08, LAST SEQUENCE UPDATE)
DE COLLAPSIN.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 94006554.
RA LEO Y., RAIBLE D., RAPER J.A.;
RT "Collapsin: a protein in brain that induces the collapse and
paralysis of neuronal growth cones";
RL CELL 75:217-227(1993).
DR EMBL; U02528; G410079; -.
DR PFAM; PF00047; Ig; 1.
SQ SEQUENCE 772 AA; 88867 MW; BC6CB9E8 CRC32;

Query Match 16.5%; Score 681; DB 13; Length 772;
Best Local Similarity 31.5%; Pred. No. 5.96e-139;
Matches 146; Conservative 119; Mismatches 150; Indels 48; Gaps 38;

Db 166 RGKSPYDPKLLTASLLVDGELYSGTA-ADFMGRDFAIFRTLGHHPHPIRTEQHDHSLW 224
Qy 135 RGYAPFSPDENSLVLFEGDEVISTIRKQYNGK-IPFRIRIGSELYTS-DIV-MQNP 190
Db 225 RFTSAHLIPSDNPEDDKIYFFRENADIGHTGKATHARIGQICKNDGFGHRS-LVNK 283
Qy 191 QFIKATIVHQ--DOAYDDKIYFFREDNPNKPEAPLVNSRVAQLCRGOGGESSLSVKW 249
Db 284 TTFKARLICSVPNGIDTHFDELQDVLM-NSKDP-KNPIYVGVFTTSSNIFKGSVC 341
Qy 250 NTFKAMLVCS--DA-ATNKNFNRLQDVFLLPDPGQWRDTRVYGVFSNPNW-Y--SVC 303
Db 342 MYSMTDVRVRLVFLGAPYHRDGNPQWVYQGRVYPRPGTCTPCTFGFSDTKDLPDEVIT 401
Qy 304 VYSLGDIKDVFTS-SLK-G-YH--S--S-LPNRPGKCLPDQ-QPI-PTETP--QV-A 349
Db 402 FARSHPAMYAPVPINS-RPIMIKTDVDYQFTQIVVDVDAEDGQ-YDVMFIDTIGTVL 459
Qy 350 -DR-HPEVAQRVEPMGPKTLPFHSK--YHYQKVAVHRMQASHGETFHVLYLTDRGTI 405
Db 460 KVSIPKETHLEELVEEVLLEEMTVPREPTVLSAMKISTKQOQLXIGSATGVSOLPLHRC 519
Qy 406 KVE-PGEQHSFAFNIEM-IQPFRAAAIQTMSLDAERKLYVSSQWESVQVPLDCEV 463
Db 520 YGRACACCLARDPYCAMDSSCSRYPFTAKRRTRRQDIRNGDPLTHCSDLQHHNPSGQ 579
Qy 464 YGGCHCLMSRDPYCGWDGRCISYSSR-RSVL-QSINPAEPHKECPN-PKDKAP-- 518
Db 580 TLEBKIIYGVENSSTFLECPKSORAIVY-WQPOKQNDHKEV 621
Qy 519 -LQ-KVSLA-PNSRYLLSC-PMESRHATYSW-HKENVEQSC 556

RESULT 13
ID O88632 PRELIMINARY; PRT; 785 AA.
AC O88632;
DT 01-NOV-1998 (TREMBLREL, 08, CREATED)
DT 01-NOV-1998 (TREMBLREL, 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL, 08, LAST SEQUENCE UPDATE)
DE SEMAPHORIN IV ISOFORM B.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ECKHARDT F., MEYERHANS A.;
RT "Molecular cloning and expression pattern of a murine semaphorin

QY 254 KAMLYCS---DAATKNENRLQDVELLPDPSCQWEDTRVYGVFSNP-WNY--SAYCVISL 307
Db 377 ADIRWVNGPFAHKEGPNYQWMPFSGKMPYPRPGTCPCGGTFTPSMKSTKDYDPDEVINEMR 436
QY 308 GDIDKVFRTS-SLK-G--YH-----SS-LPNRPGKCLPDQ-QP-I-PTETF--QVAD--R 351
Db 437 SHPLMYQAVYPLORRPLVVRT-GAPYRLTTIAVDQVDAGDG-RYEVLFGLGTDRGTQKVI 494
QY 352 -HPEVAQRVEPMG--PLKTPLFHSKYHYQKVAVHEMQASHGETFHVLYLTTRDGTIHKV 408
Db 495 VLPKDDQEMEELMLEEVEVFDPAPVKTMTISSKRQOLYVASAVGVTHLSLHRCOAYGAA 554
QY 409 E-PGEQHSFAFNEIOPFRRAAIIQTMSLDAERRKLYVSSQWEVSQVPLDLCEVYGG 467
Db 555 CADCCCLARDPYCWDGQACSRYTASKRRRRRQDVRHGNPIRQCRGFNSNANKNAVESVQ 614
QY 468 CHGCLMSRDPYCGWDQGRGIS-IYSSER-SVLQISINPAEPHKECP--NPKPDKAPLQKVS 523
Db 615 YGVAGSAAFLECPQPRSPQATYKW 537
QY 524 LA-PNSRYLLSCPMESRHATYSW 545

Search completed: Thu Jul 8 20:04:48 1999
Job time : 54 secs.